Investigations into a bHLH code for *Caenorhabditis elegans* somatic gonad regulatory cell fate and function

Hana Elisabeth Littleford

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

Abstract

Investigations into a bHLH code for *Caenorhabditis elegans* somatic gonad regulatory cell fate and function

Hana Elisabeth Littleford

The *Caenorhabditis elegans* somatic gonad is patterned by the activity of regulatory cell types, which govern its morphology, serve as the germline niche, and pattern its connection to the outside. All regulatory cell types are specified by activity of the basic helix-loop-helix gene *hlh-2/*E/Daughterless, and differences in how functions are assigned between the regulatory cells in males and hermaphrodites lead directly to their sexually-dimorphic gonads. Here, I present evidence that a code of bHLH genes function together with *hlh-2* to promote the specification and function of each regulatory cell type except for the hermaphrodite anchor cell, which is specified by HLH-2 activity alone. Each regulatory cell type expresses an overlapping but distinct set of bHLH genes, which we find are required for its specification and associated functions. Notably, ectopic expression of regulatory cell bHLH complements are sufficient to transform cells with anchor cell potential into the expected regulatory cell, albeit transiently, suggesting that they are master regulators of regulatory cell fate.

As all nematode species pattern their gonads through cognate regulatory cells and bHLH genes are highly conserved, we hypothesized that a similar bHLH code might function in specifying the regulatory cells of other species. In some nematode species the anchor cell, which remains stationary in *C. elegans*, is able to migrate. In *C. elegans*, the bHLH gene *hlh-12* is necessary for proper migration of hermaphrodite distal tip cells and male linker cell, the two migrating regulatory cell types; addition of *hlh-12* to the *C. elegans* anchor cell causes it to become displaced in a manner dependent on the endogenous hermaphrodite distal tip cell and male linker

cell machinery, suggesting that the anchor cell gains the ability to migrate with the addition of *hlh-12*. We thus hypothesized that ectopic expression of an *hlh-12* ortholog in these species might have led them to evolve migrating anchor cells. However, phylogenetic analysis of the bHLH genes of several other species, including the ones with migratory anchor cells, suggests that *hlh-12* may be novel to the *Caenorhabditis* genus and does not have orthologs in the species with migrating anchor cells, raising the possibility that either these species use another bHLH gene for migration or that their regulatory cells are specified in a bHLH-independent manner.

Table of Contents

List of Figures and Tables

Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Acknowledgments

I would not have been able to complete any of the work in this thesis without the support and guidance of my mentor, Dr. Iva Greenwald. I came to Columbia in the hopes of working with her, and my time in her lab was interesting, challenging, and always exciting. She is both a brilliant scientist and a caring mentor, and our traditional Friday evening office chats always advanced my understanding of my work and science in general. I am immensely grateful to have had the chance to work with her for the last five years, and have grown and matured as a scientist as a direct result of her mentorship.

Members of the Greenwald lab, past and present, have been without exception kind, supportive, and exceptionally generous with their time and knowledge. I am particularly indebted to Maria Sallee, who first discovered the bHLH code and laid the foundation for all of the work presented here. My rotation mentor Dr. Claudia Tenen facilitated an incredible introduction to the Greenwald lab, and Dr. Jessica Chan was the best bay mate I could ask for. Dr. Claire de la Cova was always available for coffee and to lend a sympathetic ear when experiments were not working, and Dr. Michelle Attner was a bottomless fountain of anchor cell expertise. I was excited when Sarah Finkelstein began studying inputs into the bHLH code—collaborating with her has been wonderful, and I am excited to see where she takes the code next. In addition, I had the pleasure of mentoring Henry Kim during his rotation, during which he helped construct several of the strains vital to Chapters 3 and 4. Justin Benavidez was an expert on all things HLH-2, Justin Shaffer offered unparalleled cloning advice, and Dr. Dan Shaye, Dr. Yuting Deng, Dr. Ryan Underwood, Dr. Katherine Luo, Catherine O'Keeffe, and Dr. Julia Wittes have all offered support, advice, reagents, or a combination on many occasions. Finally, the Greenwald lab would not run without its incredible lab manager Gleniza Gomez, who truly made all of my experiments possible.

My committee members were an invaluable resource in helping to guide my project through many different iterations. Dr. Oliver Hobert, Dr. Daniel Kalderon, and Dr. Richard Mann served on my qualifying committee; I would like to thank Dr. Kalderon and Dr. Hobert in particular for continuing on my committee throughout my time, and for offering helpful suggestions at committee meetings that often cast my project in a new and exciting light. Their guidance has helped my project take its final shape, for which I am very grateful. In addition, I thank Dr. Xantha Karp and Dr. David Fitch for agreeing to be on my final thesis committee.

I am grateful to have been part of the *C. elegans* community of both Columbia and New York City, because of all the support and advice I have received. Dr. Fitch and Dr. Karin Kiontke kindly offered advice on the work done in Chapter 4, which could not have been completed without their help, and joint lab meetings with the Hobert and Chalfie labs were a great chance to get outside perspectives on my work.

I could not have completed any part of this work without support from family and friends. My parents, Alan Littleford and Lang Anh Pham, have always encouraged me to do what I love, and their unfailing support has been the foundation for everything I have achieved. My brother Timon's visits were a welcome distraction from the lab. While our get-togethers were far too few, my friends were always there to offer distraction or support as needed. Closer to home, Dr. Rebecca Cox made my time in New York City fun and was always there when I needed anything, scientific or otherwise.

Finally, my fiancé Ji-Sup Yang has offered nothing but support, understanding, and love throughout the last six years. Whether it was helping me sort out technical difficulties with my experiments or calming me down when I was too stressed out to think, he made every part of this process possible. I am immensely grateful for his presence in my life.

Chapter 1: Introduction

Introduction

In every animal, cells must specify their proper fates in order to execute necessary functions for development. In this thesis, I used the model system *Caenorhabditis elegans* to study cell fate determination because of its almost-invariant lineage of cell divisions and fates, amenability to genetic manipulations, and the ability to view developmental events *in vivo*. In particular, I have addressed how specific bHLH transcription factor complements contribute to the fates and functions of crucial cells that regulate *C. elegans* gonad development, and the consequences of altering their bHLH complements on their fates. In addition, I searched for orthologs of the key *C. elegans* bHLH transcription factors in other species to determine if they might play similar roles in gonad patterning.

1. *C. elegans* gonadogenesis

The adult gonad is one of the largest organs in both the male and the hermaphrodite worm. It contains the germline, which is encapsulated by the somatic gonad. The somatic gonad stimulates the dividing germline starting in the earliest developmental (L1) stage, providing a constant supply of gametes through adulthood (Kimble & Hirsh, 1979). In addition, the somatic gonad forms the structures necessary for exit of the gametes or embryo: the uterus and uterine-vulval connection in the hermaphrodite, and the cloaca in the male (Kimble & Hirsh, 1979; Sulston & Horvitz, 1977). At hatching, males and hermaphrodites have gonads containing just four cells, named Z1-Z4; of these, Z2 and Z3 will give rise to the germline, while the entire somatic gonad will come from Z1 and Z4 (Figure 1.1A, Kimble & Hirsh, 1979).

1.1 Hermaphrodite/female somatic gonad development and female regulatory cells (see Figure 1.1)

C. elegans has two sexes, the male and the hermaphrodite. However, hermaphrodites are hermaphroditic only because they transiently make sperm; their somatic structures resemble those of female nematodes from other species, and thus they are considered somatic females (reviewed in Zarkower, 2006).

The adult hermaphrodite has a bilaterally-symmetric, U-shaped gonad, with reflexed arms extending anterior and posterior and a vulva located at the midbody (Figure 1.1D). It develops from the two gonad precursor cells Z1 and Z4, which divide to form twelve cells by the late L1 stage (Figure 1.1A, B). By mid L2, three of the cells have specified into female regulatory cell types: two hermaphrodite distal tip cells (hDTCs), and a single anchor cell (AC, Figure 1.1C). I will refer to the stage at which the regulatory cells have just specified as the somatic gonad primordium stage for the rest of this work. The two hDTCs will go on to lead extension of the two gonad arms and serve as the germline niche, while the AC remains fixed at the midbody where it is born and patterns the vulva and the ventral uterus (Kimble & Hirsh, 1979; Sternberg, 2005; Newman et al., 1995).

1.2 Male somatic gonad development and male regulatory cells (see Figure 1.1)

Like the hermaphrodite gonad, the male somatic gonad develops from Z1 and Z4. However, in the male these two gonad precursor cells divide to form only 10 cells by the somatic gonad primordium stage, compared to the 12-cell hermaphrodite somatic gonad primordium (Figure 1.1B). The difference lies in the distal daughters of Z1 and Z4: in the hermaphrodite, these divide once more to form the hDTCs and their sisters, whereas in the male the distal daughters

specify directly as male distal tip cells (mDTCs, Figure 1.1B). In addition, there is a global rearrangement of cells in the male gonad, starting at the four-cell stage when the anterior Z1-derived mDTC migrates to the posterior to join the Z4-derived mDTC (Kimble & Hirsh, 1979). This results in two neighboring mDTCs at the distal end of the gonad, where they remain fixed and serve as the germline niche. Extension of the single male gonad arm is led by the linker cell (LC), which specifies at the proximal end and migrates to form the J-shaped gonad terminating at the cloaca at the tail in the adult (Figure 1.1D).

1.3 Specification of the AC: the AC/VU decision (see Figure 1.2)

The AC and LC are unique in their respective gonads because they specify using lateral signaling facilitated by LIN-12/Notch. In both animals, two neighboring proximal cells will resolve a decision where one becomes an AC and the other a ventral uterine cell (VU) in the hermaphrodite, and one becomes an LC and the other a vas deferens cell (VD) in the male (Kimble & Hirsh, 1979). Of the two events, the most work has been done on AC specification. The AC specification event happens between two cells named Z1.ppp and Z4.aaa, one of which will always specify as an AC and the other of which will always specify as a VU. Both cells are born expressing both LIN-12/Notch and its ligand LAG-2/Delta; over time, as the decision resolves, one cell will express only LIN-12 and the other only LAG-2. The LAG-2-expressing cell will become the AC, and the LIN-12-expressing cell the VU (Figure 1.2A, Greenwald et al., 1983; Lambie & Kimble, 1991).

The basic helix-loop-helix (bHLH) transcription factor *hlh-2*/E2A/Daughterless is required at several stages for AC competence and specification. Initially, all four daughters of Z1.pp and Z4.aa are born with the ability to become the AC; the distal cells, termed the β cells, will always

become VUs in wild-type conditions, while the proximal cells or α cells undergo the AC/VU decision (Kimble & Hirsh, 1979; Seydoux & Greenwald, 1999). Xantha Karp, a previous member of the Greenwald lab, found that RNA interference (RNAi) against *hlh-2* at different stages leads to diffeent AC defects: if done early in L1, worms specify 0 ACs, while if RNAi is done in late L1 they specify 2 ACs (Figure 1.2A, Karp & Greenwald, 2004). This suggests that HLH-2 acts first as a "pro-AC" factor that gives the competence for AC fate, and later plays a more direct role in the specification event through direct transcriptional regulation of *lag-2* (Karp & Greenwald, 2003).

Initially, which cell became the AC and which became the VU was thought to be a purely stochastic event, as in wild-type conditions the AC forms from Z1.ppp 50% of the time and from Z4.aaa 50% of the time (Kimble & Hirsh, 1979). However, Xantha Karp found a bias in which cell was fated to become the AC. As the two cells involved in the decision are not sister cells, they are born at different times; Xantha found that the cell born first was biased towards becoming the VU (Karp & Greenwald, 2003). Recently, more work done in the Greenwald lab has uncovered a potential mechanism for the apparent birth order bias (Attner et al., 2019). In this model, there are two initial differences between the cells involved in the decision: the timing of their birth, which can range from under ten minutes to over two hours apart, and expression of HLH-2 in the parent cells (Figure 2.2B). HLH-2 is expressed in both parent cells, but comes on first in the parent of the VU; after HLH-2 expression onset, HLH-2 then activates LIN-12 and LAG-2 expression, which continues in both daughter cells initially after they divide. Due to the difference in HLH-2 expression timing, one daughter cell is likely born with an "edge" in LIN-12-activated signaling granted by *lin-12* transcript, HLH-2 protein, or some chromatin pioneer factors. LIN-12 on this first-born cell can be activated by LAG-2 ligand on either the second-born cell or its parent to

amplify these initial differences, starting the feedback mechanism in which LIN-12 signaling activity autoregulates its own expression but turns off *hlh-2* expression and thus *lag-2* transcription. Thus, the cell descended from the lineage with earlier HLH-2 expression is biased strongly towards adopting the VU fate. As HLH-2 expression invariably begins approximately two hours after the parent cell is born and does not seem to be affected by cell cycle length, it ties also into the birth-order bias: in animals where Z1.ppp and Z4.aaa are born greater than 24 minutes apart, the increased difference in HLH-2 timing onset between the parents leads to a buildup of more LIN-12 activity in the first-born daughter, and thus the fist-born daughter invariably becomes the VU. However, in animals where the two cells are born fewer than 24 minutes apart, the decision is predicted not by birth order but by which parent cell first expressed HLH-2 (Attner et al., 2019).

1.4 *hlh-2* is required for the specification of the LC and hDTCs

As mentioned above, the bHLH gene *hlh-2*/E2A/Daughterless is necessary for AC competence and specification. *hlh-2* is also necessary for the proper specification of the hDTCs and LC. Xantha Karp found that hDTCs did not extend in animals with *hlh-2* RNAi treatment beginning in early L1, suggesting that *hlh-2* is required for their specification and/or migration (Karp & Greenwald, 2004). Chesney et al. (2009) performed *hlh-2* RNAi on a hypomorphic *hlh-2* allele to further explore *hlh-2*'s role in regulatory cell specification. They confirmed that *hlh-2(lf) + RNAi* animals had hDTCs that did not migrate or express the fate marker *lag-2*, and found in addition that *hlh-2(lf) + RNAi* animals had LCs which again did not migrate or express *lag-2*. Although they did not find a requirement for *hlh-2* in mDTC specification, they did find that the mitotic germline region in males was smaller in *hlh-2(lf) + RNAi* animals, suggesting that *hlh-2* plays at least a functional role in the mDTCs. Together, their results when combined with Xantha's

RNAi studies suggests that *hlh-2* is required for specification or function in all regulatory cell types.

2. Functions of the somatic gonad regulatory cells

For a list of markers used for each cell fate and function, see Table 1

2.1 Patterning of the vulva and cloaca connections (see Figure 1.3)

Patterning of the connection that allows gametes to exit the gonad is done by the proximal regulatory cell types in each sex. In the hermaphrodite, the AC is responsible for patterning the developing vulva and ventral uterus, and also facilitates the connection between the two. First, the AC directly induces the fate of the hypodermal vulval precursor cells (VPCs) by sending an EGF signal which is received by P6.p, the closest adjacent VPC (Figure 1.3A, Sternberg, 2005). The *C. elegans* EGF ortholog is *lin-3,* which is a direct *hlh-2* target (Sallee & Greenwald, 2015). This EGF signal leads to a signaling cascade that results in P6.p adopting the primary fate and using LIN-12/Notch signaling to specify its two neighbors, P5.p and P7.p, as secondary fate. Descendants of the cells with primary and secondary fate will later form the vulva. The remaining VPCs, P3.p (not shown in Figure 1.3A), P4.p, and P8.p, adopt a tertiary fate in the absence of EGF or Notch signaling, and their daughters later fuse with the hypodermis (reviewed in Sternberg, 2005). In the absence of an AC, P6.p does not specify as primary and does not induce its neighbors to specify as secondary, meaning that all cells adopt tertiary fate in the absence of signaling and fuse with the hypodermis, leading to a complete absence of the vulva (reviewed in Sternberg, 2005).

Beginning in L3, the AC will dissolve the basement membrane that separates it from the VPC descendants, and send down invadopodia to directly contact the vulval cells (Figure 1.3B).

This process is required to induce their invagination, and later to initiate a proper connection between the uterus and the vulva (Hagedorn & Sherwood, 2011; Newman & Sternberg, 1996; Sherwood & Sternberg, 2003). AC invasion is facilitated by the transcription factor *fos-1*, which is required for proper invasion and controls expression of a number of effector genes required for dissolving the basement membrane and inducing later fusion of the AC (Sapir et al., 2007; Sherwood et al., 2005). As with induction, *hlh-2* regulates key genes required for invasion, including *fos-1* and the direct invasion effector *cdh-3* (Medwig-Kinney et al., 2020; Schindler & Sherwood, 2011).

The AC is also responsible for patterning the developing ventral uterus, and facilitating the connection between the uterus and the vulva. In L3, the three VU cells will divide to form a total of 32 nuclei in the fully-formed ventral uterus. The VU granddaughters will specify as either π or ρ cells, which have different morphologies and produce different numbers of descendants (Kimble & Hirsh, 1979; Newman et al., 1995). π vs. ρ cell fate is also determined by LIN-12-mediated lateral signaling events coordinated by the AC: both eventual cell types express LIN-12, but π-fated cells, located next to the LAG-2-expressing AC, receive the ligand signal and the more distant ρ-fated cells do not (Figure 1.3C, Newman et al., 1995). As *hlh-2* directly controls *lag-2* expression in the AC (Karp & Greenwald, 2003), it is required for proper ventral uterine patterning via this specification event.

Later in L4, eight of the π cell descendants fuse to form the syncytial uterine seam cell (utse), which sits around and above the developing vulva and holds the uterus in place by connecting to the hypodermis (Gupta, Hanna-Rose, & Sternberg, 2012; Newman & Sternberg, 1996; Sternberg, 2005). At this stage, the AC sits directly above the developing vulval pore, blocking connection between the vulva and the uterus. The AC will eventually fuse with the utse

by late L4 in a process regulated again by *hlh-2*, leaving only a thin layer of cytoplasm separating the uterus and the vulva (Figure 1.3D, Sapir et al., 2007; reviewed in Gupta et al., 2012). This fragment of cytoplasm will break when the first embryo is laid, resulting in a fully functioning vulva.

Like the AC, the LC facilitates the connection of the developing gonad to the opening, which is the cloaca in males. Unlike the AC, however, the LC does not induce cell fate or regulate patterning of the developing cloaca. Instead, the cloaca develops in the tail while the LC is still leading gonad migration; once the LC reaches the forming cloaca in L4 it undergoes cell death and is engulfed by the neighboring epithelial cells (Sulston & Horvitz, 1977). This is a fairly unique event in *C. elegans*, as LC death does not require any part of the core apoptotic pathway or even the presence of the engulfing epithelial cells (Abraham, Lu, & Shaham, 2007; reviewed in Malin & Shaham, 2015). Instead, cues from the Wnt and heterochronic pathways promote LC death, which is dependent on activity of the transcription factor *hsf-1*. Interestingly, *hsf-1* ordinarily protects against cell death caused by heat or other stressors; in the case of the LC, it instead appears to coordinate cell death by promoting expression of LET-70/E2 ligase and recruiting the proteasome (Kinet et al., 2016). The engulfment of the LC corpse by its neighboring cells is also atypical and does not require the usual genes involved in apoptotic cell clearance (Kutscher, Keil, & Shaham, 2018). Engulfment of the cell corpse by the neighboring epithelial cells facilitates fusion between the gonad and the cloaca, allowing for proper sperm exit (Abraham et al., 2007; Malin & Shaham, 2015).

2.2 Migration of the hDTCs and LC (see Figure 1.4)

The gonad shape is determined by activity of the migrating cell types, which differ between the sexes: in hermaphrodites it is the two hDTCs and in males it is the single LC, which leads directly to the different gonad morphologies of the two sexes. I will refer to the hDTCs and LC jointly as the leading regulatory cell types. Throughout the migration process, both the leading regulatory cell types must coordinate inputs for proper initiation, turning, pathfinding, and cessation to govern overall gonad shape. While the two cell types do have the same basic function, they have very different lineal origins and positional identities, and slightly different migration paths (Figure 1.4). Both lead outgrowth beginning in early L2, shortly after their respective specification events in late L1. In hermaphrodites, the two hDTCs extend in opposite directions on the A/P axis along the ventral side of the animal, before turning first dorsally and then proximally by the end of L3 and terminating just dorsal to the vulva in L4 (Figure 1.4A). In males, the LC extends first anterior, and makes a dorsal and posterior turn by the end of L2; it extends along the top of the distal end of the gonad throughout L3, before turning ventral and posterior once more and migrating along the ventral side of the animal to join with the cloaca at the tail in L4 (Figure 1.4B). It is interesting to note that the LC starts turning a full larval stage before the hDTCs, completing a dorsal and posterior turn by the end of L2 while the hDTCs complete their dorsal and proximal turns by the end of L3 (reviewed in Sherwood & Plastino, 2018; Wong & Schwarzbauer, 2012).

Proper gonad arm extension requires both correct pathfinding and the ability of the regulatory cell to physically migrate. In the hermaphrodite, the bHLH transcription factor *hlh-12* is required for the migration of the hDTCs: it is expressed in the hDTCs shortly after their specification, and directly activates targets that help with physical migration (Blelloch & Kimble,

1999; Meighan & Schwarzbauer, 2007; Tamai & Nishiwaki, 2007a). For example, expression of the ADAMTS metalloprotease *gon-1* helps to dissolve the extracellular matrix, enabling passage of the hDTC. HLH-12 also activates expression of the α-integrin *ina-1*, which is thought to hook into the basement membrane and help pull the migrating cell forward along the ventral and dorsal surfaces (Meighan & Schwarzbauer, 2007; reviewed in Wong & Schwarzbauer, 2012). Pathfinding requires a number of different genes at each step in hDTC migration, many of which are also associated with the basement membrane. For example, the ADAM protease *mig-17* is predicted to modify collagens to ensure proper direction, while the α-integrin *pat-2* is required for proper pathfinding along the dorsal body wall (Meighan & Schwarzbauer, 2007; Sherwood & Plastino, 2018). In addition, diffusible cues also help orient the migrating hDTC: the netrin *unc-5* is required for the dorsal turn, and Wnt pathway members are required for proper A/P migration (Cabello et al., 2010; Hedgecock et al., 1990; Sherwood & Plastino, 2018).

LC and hDTC migration share many similarities. LC migration is also enabled by *hlh-12*, which acts again through both *gon-1* and *ina-1* (Blelloch & Kimble, 1999; Meighan & Schwarzbauer, 2007; Tamai & Nishiwaki, 2007a). In addition, it shares pathfinding genes with the hDTCs. Notably, the netrin *unc-5* is required for proper turning in both sexes, but it regulates the entire dorsal-posterior turn that the LC makes in late L2, while in hermaphrodites it regulates only the dorsal hDTC turns in L3 (Kato & Sternberg, 2009a; Su et al., 2000). This difference in the roles that the same pathfinding genes play in hermaphrodite and male gonads is a reoccurring theme. *daf-12*, a nuclear hormone receptor that governs dauer diapause, fat metabolism, and lifespan, also regulates pathfinding in both males and hermaphrodites: various mutant alleles show defects in pathfinding beginning in the L3 and L4 migration stages (Antebi et al., 1998; Motola et al., 2006). Interestingly, hermaphrodites with a hypomorphic *daf-12* allele containing a predicted

deletion of the ligand-binding domain fail to make the dorsal and proximal turns ~92% of the time, while ~85% of males with this allele make the dorsal and posterior turn properly (Antebi et al., 1998). These same animals also showed ~85% loss of the later ventral and posterior turns in the male, underscoring the differences in timing between the migrations of these two different cell types.

As discussed above, the LC employs many of the same mechanisms for the physical act of cell migration as the hDTC, as it too expresses *hlh-12, gon-1,* and *ina-1* (Blelloch & Kimble, 1999; Meighan & Schwarzbauer, 2007; Tamai & Nishiwaki, 2007a). However, the genetic control of its migration appears more complicated than that of the hDTC. Kato et al (2010) propose a model in which aspects of the LC's basal migration and pathfinding programs are modified by stage-specific regulators, transcription factors which control the programs required for turning in L2 and LC morphogenesis and increased migration speed in L4. One of these stage-specific regulators is *nhr-67/tailless*, which they find regulates migration from early L3 to mid L4 and is necessary for downregulating *unc-5* expression (Kato & Sternberg, 2009a). Interestingly, *nhr-67* is not expressed in the hDTCs, but it is expressed in the non-migrating AC where it helps regulate multiple stages of uterine development (Verghese et al., 2011). This shared expression of *nhr-67* in the non-migrating AC and migrating LC could be due to their similar lineal origins or proximal identity, further demonstrating the importance of these factors in understanding regulatory cell specification and function.

2.3 The hDTC and mDTC germline niche function (see Figure 1.5)

The gonad of each sex contains two distal tip cells: the hDTCs in the hermaphrodite, and the mDTCs in the male. Both cell types are specified by late L1, and serve as the germline niche

(Figure 1.5, Kimble & Hirsh, 1979). The *C. elegans* germline originates from the primordial germ cells Z2 and Z3, which generate the germline stem cells (GSCs) that remain at the distal end(s) of the gonad, adjacent to the DTCs. The GSCs express GLP-1/Notch, which is activated by a number of DSL ligands expressed by the distal tip cells to promote GSC mitosis (Austin & Kimble, 1987). Thus, the distal tip cells of both sexes act as the germline niche by enabling mitosis of the stem cell progenitor pool, providing a constant supply of germline throughout the worm's life cycle.

The hDTCs have a striking morphology, with long projections that can extend for a distance of ten GSC diameters or more (Hall et al., 1999); as daughters of the dividing cells are pushed out of the reach of the distal tip cell projections, they lose contact with the ligands and enter meiosis (Figure 1.5A, reviewed in Byrd & Kimble, 2009). In *C. elegans*, there are 10 predicted DSL ligands: *lag-2, apx-1, arg-1,* and *dsl-1-7* (N. Chen & Greenwald, 2004). The hDTCs express both *lag-2* and *apx-1*, which are functionally redundant for germline stimulation (Byrd & Kimble, 2009; Nadarajan et al., 2009).

Unlike the hDTCs, the mDTCs in the adult do not have projections, though the Greenwald lab has found that the WT mDTCs do display projections towards the germline during larval development (Figure 1.5B, data not shown; Crittenden et al., 2019; Sallee et al., 2017). There are also sexual dimorphisms in the ligands expressed, as the mDTCs do not express *lag-2* but do express *arg-1*, which is not expressed in the hDTCs (Salle et al., 2017). However, *apx-1* is expressed in both DTC types. In addition, the presence of both mDTCs at the distal end of the gonad may cause some other differences in niche architecture between the sexes: the male niche has a bulge at the very distal end that the hermaphrodite niche lacks, and the asymmetric positioning of the mDTC cell bodies means that a significant portion of GSCs have no direct contact with the niche (Figure 1.5; Crittenden et al., 2019).

3. bHLH transcription factors

3.1 Overview and classification systems

Basic helix-loop-helix (bHLH) transcription factors are a large, conserved family of transcription factors found in organisms ranging from yeasts to humans. They play crucial roles in a number of developmental processes, including neurogenesis, myogenesis, hematopoiesis, sex determination, and more (Massari & Murre, 2000). In humans, dysregulation in their function is linked to cancers such as leukemias and lymphomas, especially T-cell lymphomas (Bain et al., 1997). In model organisms like *Drosophila* and *C. elegans*, bHLH factors are known for their important role in promoting lineage commitment and cell fate specification (Massari & Murre, 2000).

bHLH proteins are defined by the presence of two adjacent domains: the basic region, which is required for DNA binding, and the helix-loop-helix domain, which is required for dimerization (Figure 1.6A, Ferré-D'Amaré et al., 1993; Massari & Murre, 2000). All bHLH proteins are obligate dimers, and must interact through their helix-loop-helix domains in order for their basic domains to contact E boxes on the DNA and promote transcription (Figure 1.6B). All E boxes have the structure CANNTG, with the basic domain of each monomer binding to half of the sequence in an interaction further stabilized by specific residues in the loop and helix 2 (Ellenberger et al., 1994). Due to differences in helix-loop-helix domain structure, some bHLH proteins are capable of forming homodimers or heterodimers, while others can only form specific heterodimers; these differences in dimerization partners means that each type of dimer recognizes a different E box sequence or sequences, leading to activation of precise targets (Murre et al., 1989; reviewed in Massari & Murre, 2000).

The bHLH family is extensive, with 42 members in *C. elegans* and 125 in humans, and individual bHLH proteins differ in their dimerization ability, E box targets, expression patterns, other interactors, and more (Baker & Brown, 2018; Grove et al., 2009; Massari & Murre, 2000). Massari and Murre (2000) developed a method for classifying bHLH proteins, dividing them into seven classes based on some of these factors. Their Class I contains proteins containing only the bHLH domains, with the ability to form either homo- or heterodimers and which have generally broad expression patterns. *C. elegans* has one Class I bHLH protein, HLH-2 (Krause et al., 1997b); humans have four E proteins and Drosophila has one, Daughterless (Caudy et al., 1988; Cronmiller & Cummings, 1993; Massari & Murre, 2000). Class II proteins are only able to form heterodimers with Class I proteins (Figure 1.6B), and have much more temporally- and/or spatially-restricted expression patterns leading to tight regulation of active heterodimer in developmental processes. There are eighteen predicted Class II bHLH proteins in *C. elegans* that can be further subdivided into groups based on orthology to fly genes like Twist, Atonal, and more. In addition, there are a few Class II bHLH genes in *C. elegans* that are regarded as orphan genes due to their lack of resemblance to any other known bHLH proteins (Grove et al., 2009.; Ledent, et al., 2002).

This work focuses exclusively on Class I and Class II proteins as defined by Massari and Murre, and all non-*hlh-2 C. elegans* bHLH genes mentioned here are Class II unless specified otherwise. For a further description of this class system and other phylogeny-based methods of classification, as well as a discussion of bHLH conservation overall, see Chapter 4.

3.2 Examples of bHLH-induced fate specification in *C. elegans*

In *C. elegans*, bHLH genes have been studied in contexts other than the gonad, and in particular studies of bHLH genes in neuronal and muscle lineage specification provide examples

relevant to the work I will present in Chapter 3. As in *Drosophila* and mammals, *C. elegans* Class II genes are important for proneural potential and neural fate, and for muscle development (reviewed in Massari & Murre, 2000). In this capacity, they have been observed to act earlier in the lineage to promote neuronal identity of cells later, or they can promote cell fate specification in the differentiating neuron itself.

A simple example in which bHLH factors act at the level of the parent cells to specify cell fate is given by the specification of the *C. elegans* motor neuron MI. This neuron is formed after the asymmetric division of a parent blastomere to generate an epithelial and a neuronal daughter (Sulston & Horvitz, 1977). The bHLH gene *ngn-1*/*neurogenin* is required for this asymmetric division, as *ngn-1(0)* animals have two epithelial cells and no MI neurons (Nakano et al., 2010). In this instance, a signaling cascade beginning three generations removed from MI and its sister leads to expression of NGN-1 and HLH-2 in the MI parent, which then leads to proper MI specification in the next generation (Nakano et al., 2010).

A more complex example is that of *lin-32*/*atonal* in the *C. elegans* male tail neuron lineage, as *lin-32* is required at multiple steps in precise lineages for proper neuronal specification (Figure 1.6B, Portman & Emmons, 2000). The male tail, a copulation structure, contains multiple sensory ray neurons required for sensing the hermaphrodite vulva. These ray neurons are formed from three sets of asymmetric cell divisions: a ray precursor gives rise to a hypodermal cell and a ray neuroblast, the latter of which divides twice more to give rise to two neurons and a neuron support cell (Figure 1.6C, see Portman & Emmons, 2000). Studies disagree as to whether or not LIN-32 expression begins at the ray precursor or ray neuroblast stage (Miller & Portman, 2011; Portman & Emmons, 2000a). However, LIN-32 is expressed in the anterior daughter of the ray neuroblast, though not the A-type neuronal granddaughter; it is not expressed in the posterior ray neuroblast

daughter that gives rise to the B-type neuronal granddaughter (Portman & Emmons, 2000a). *lin-32(0)* animals still have some ray neurons present, though at a very reduced number, suggesting that *lin-32* is not completely required for ray neuron fate. However, overriding the wild-type asymmetric expression in the lineage via heat-shock expression results in the loss of B type neurons from the ordinarily *lin-32*(-) lineage, as well as the generation of some ectopic A type neurons (Portman & Emmons, 2000a). This suggests that *lin-32* is required at multiple levels in very precise lineages for proper patterning of the male tail ray neurons overall.

bHLH proteins have also been found to act in the end cells of the lineage to specify their fates. One example of this is *hlh-4'*s activity in the ADL nociceptor sensory neuron (Figure 1.6D, (Masoudi et al., 2018). The ADL lineage is initially given proneural fate by the activity of *hlh-14/*Achaete-Scute; *hlh-4*/Achaete-Scute is expressed only in ADL and its apoptotic sister cell after the parental division. In ADL, *hlh-4* regulates both morphology and the expression of a number of genes required for chemoreceptor expression and chemosensory function, suggesting that it is the terminal selector for ADL fate and function (Masoudi et al., 2018). Another example is the role of another Achaete-Scute homologue, *hlh-3*. While *hlh-3* expression is seen in all neuronal precursor lineages, consistent with a proneural role *a la lin-32* as described above, *hlh-3* may also have a role in terminal differentiation of some neurons, as expression is retained in differentiating HSN and ventral cord neurons. In particular, HSN neurons in *hlh-3(0)* animals lack serotonin biosynthesis gene expression and have axonal morphology defects (Doonan et al., 2008), suggesting that they are specified improperly in the absence of HLH-3.

4. Reprogramming between cell types in *C. elegans*

4.1 Overview

In the classical model, cells start out with the ability to adopt any fate; as they divide, their ability to adopt fates narrows, until they finally specify as one cell fate post-mitotically and retain that fate for the rest of their life (Waddington, 1957). However, more recent work has shown that cellular plasticity is much more ubiquitous and persistent than previously thought, and that even terminally-differentiated cells can be stimulated to adopt new fates *in vitro* or *in vivo*. The most striking example is that of induced pluripotent stem cells (iPSCs), in which any somatic cell can be treated with a specific cocktail of transcription factors to revert back to pluripotency (Takahashi & Yamanaka, 2006); cells can also be directly reprogrammed *in vitro* from one terminal fate to another, skipping the pluripotency stage altogether, or even undergo reprogramming during the course of normal development (reviewed in Rothman & Jarriault, 2019). Understanding how cells can be induced to adopt different fates at any point in their life has great potential for regenerative therapies.

The term "reprogramming" itself covers a wide variety of cellular events. Reprogramming can be achieved through experimental manipulation, or as a part of natural development or response to environmental changes or trauma, and can occur between cells at various stages of fate commitment. To date, cells have been reprogrammed artificially from terminal differentiation to multipotency, from lineage progenitors to a different terminal lineage, and directly from one terminal identity to another in *C. elegans* alone (Euling & Ambros, 1996; Fukushige & Krause, 2005; Riddle et al., 2016). In addition, there are several natural reprogramming events, including regeneration from injury in amphibians and an epidermal to neuron transformation during development in *C. elegans* (Jarriault et al., 2008; reviewed in Henry & Tsonis, 2010). While these

different events have precise names (for review see Rothman & Jarriault, 2019), I have used the general "reprogramming" here to address them all for the sake of simplicity, with "direct reprogramming" used to indicate a reprogramming event from one differentiated fate to another without passing through any intermediary stage.

While reprogramming events can take place between different stages of potency, happen *in vitro* or *in vivo*, and can occur experimentally, as part of natural development, or in response to injury, all of these events share several hallmarks. The most important is a change in overall nuclear transcription, which translates to changes in morphology and function associated with the new cell type (Rothman & Jarriault, 2019). In addition, while these events may require passage through several stages of de-differentiation and re-differentiation, reprogramming is considered permanent in its final state; for example, studies have shown that iPSCs have permanent changes in gene regulation characterized by alterations in epigenetic marks (Federation et al., 2014). Finally, some direct reprogramming events are characterized by a mixed stage, in which the cell transiently expresses transcriptional programs characteristic of both its initial and final fate (Marro et al., 2011). These hallmarks provide an important set of criteria for determining if a reprogramming event is truly occurring.

4.2 Natural reprogramming events

As mentioned above, there are several natural cases of cell reprogramming that can happen in either normal developmental or regenerative processes. For example, adult newts can regenerate their irises by direct reprogramming of nearby pigmented epithelial cells into lens cells (Kodama & Eguchi, 1995). In addition, the process of limb regeneration in newts involves a large-scale reprogramming of terminally-differentiated muscle, cartilage, and connective tissue cells at the

wound site into blastemal cells that coordinate outgrowth of the new limb, before re-differentiating into their proper types after the process is complete (reviewed in Brockes & Kumar, 2002). Though mammals cannot undergo such extensive regeneration, there is some evidence of reprogramming after injury in adult animals, such as direct reprogramming of pancreas to liver cells after depletion of copper in rats (Rao et al., 1988).

In *C. elegans*, there are several examples of natural reprogramming (reviewed in Rothman & Jarriault, 2019). The most studied such example is the Y-to-PDA transition, in which Y, a rectal cell, directly reprograms into the motor neuron PDA during late L1, a process which involves changes in cellular morphology and migration away from the rectum. This event does not require contribution of the neighboring rectal cells or even migration, as ablation of rectal cells or blocking migration did not result in a change in reprogramming. However, it does require *lin-12* activity in order to give Y the competence for reprogramming (Jarriault et al., 2008). Notably, this appears to be an example of direct reprogramming where the cell does not adopt a mixed identity as a transition point: despite several screens, there have been no alleles identified to date which cause co-expression of both Y and PDA markers (reviewed in Rothman & Jarriault, 2019). Instead, the entire process appears to take place in a series of sequential stepwise events, in which the Y cell first de-differentiates and then a coordinated set of transcription factors and histone modifiers further programs the de-differentiated cell into an invariant PDA (Richard et al., 2011). Interestingly, the transcription factors required for the initial de-differentiation of Y have mammalian orthologs which associate with the transcription factors required for iPSC formation, hinting at a possible conserved set of cellular plasticity effectors (Jarriault et al., 2008; Rothman & Jarriault, 2019).

While the Y-to-PDA involves conversion between two very different cell types, there is evidence for several other natural reprogramming events in *C. elegans* as well. Of particular interest to us is the adoption of I4 neuron fate. While most neurons arise from the AB lineage, I4 instead comes from the MS lineage, which mostly gives rise to mesodermal tissues (Sulston & Horvitz, 1977). The parent cell of I4, generally understood to have a mesodermal identity, divides asymmetrically to give rise to I4 and a pharyngeal muscle daughter. Interestingly, the parent cell, I4, and its sister muscle cell all express markers of muscle fate, including the bHLH gene *hlh-1/MyoD*. This muscle expression turns off shortly after birth in I4 only as it adopts its neuronal fate, suggesting a natural reprogramming event from muscle to neuronal fate in this one cell (Luo & Horvitz, 2017). Interestingly, the bHLH gene *hlh-3/Achaete-Scute* was found to be necessary for adoption of I4 identity. HLH-2 and HLH-3 are expressed in only I4 shortly after its birth, and in the absence of HLH-3 the cell fated to become I4 adopts the fate of its sister pharyngeal muscle cell instead (Luo & Horvitz, 2017).

4.3 Induced reprogramming events

As discussed above, there have been many recent attempts at experimentally-induced reprogramming. Here, I discuss several of note in both mammals and *C. elegans*, including some attempts using bHLH factors to induce reprogramming.

Many reprogramming events in *C. elegans* have been induced by ectopic expression of a large amount of protein, mostly from a ubiquitous heat-shock promoter. One such example comes from Tursun et al. (2011), who used a heat-shock promoter to express ectopic *che-1*, a master regulator of ASE fate. However, they were able to induce ectopic expression of the ASE fate markers *gcy-5* and *ceh-36* only in the germline, suggesting that the germline cells were particularly

susceptible to reprogramming by this factor; they found further that the germline cells had nuclear and morphology changes consistent with reprogramming into neurons. Similarly, Riddle et al. (2016) used heat-shock promoters to express the intestinal transcription factor *elt-7* in brief pulses as late as L4. Remarkably, they showed conversion of the entire hermaphrodite somatic gonad into intestine after this brief and late expression, as shown by both marker expression and electron micrographs. This conversion is of particular interest to us because it suggests that the cells in the somatic gonad are able to be reprogrammed by expression of just a single transcription factor, even well after their initial fates are specified.

Lastly, bHLH factors have been used to induce reprogramming in several different contexts. Davis et al. (1987) first reported that transfection of mouse fibroblasts with myoD cDNA transformed them into myoblasts. Later, myoD was found to induce myoblast fate in differentiated cell lines representing all three germ layers, as well as fibroblasts from chickens, rats, and humans, but not in monkey kidney cells or HeLa cells (Weintraub et al., 1989), underscoring the importance of the initial cell fate in each reprogramming context. In addition, transfection of mouse carcinoma cells with neural Class II bHLH factors including NeuroD2 and Ngn1, along with the Class I E protein E12, resulted in reprogramming into neurons as seen by morphological changes, marker expression, and the ability to conduct electrical impulses (Farah et al., 2000). This work also showed that this bHLH-induced reprogramming was directly tied to exit from the cell cycle, suggesting a potential link between exit from the cell cycle and natural bHLH-induced fate differentiation.

bHLH factors have also been used to induce cell fate reprogramming in *C. elegans* in a few separate instances. As discussed in section 3.2, the Achaete-Scute homologs *hlh-3* and *hlh-4* are necessary for specification in the specifying I4 and ADL neurons, respectively (Luo & Horvitz,

2017; Masoudi et al., 2018). When expressed ubiquitously along with *hlh-2*, *hlh-3* caused ectopic expression of pan-neuronal markers in their body wall muscles that suggested muscle-to-neuronal reprogramming (Luo & Horvitz, 2017). However, while these cells expressed neuronal markers and projections consistent with neuronal morphology, they retained the muscle nuclear morphology, suggesting a mixed overall fate (Luo & Horvitz, 2017). In addition, *hlh-4*, which directly induces wild-type ADL fate, can induce marker and/or morphological changes consistent with ADL fate when expressed in other ADL-like ciliated sensory neurons or head neurons (Masoudi et al., 2018).

5. Structure of this thesis

As discussed above, the *C. elegans* gonad in both the hermaphrodite and the male is patterned by the activity of specific regulatory cell types, whose specification and function are dependent on the activity of bHLH proteins. In addition, bHLH proteins are capable of reprogramming across cell types, and their sufficiency for regulatory cell fate led us to hypothesize that changes in their expression during evolution might underlie the changes in regulatory cell function seen in other nematode species.

In **Chapter 2,** I show that a "code" of Class II bHLH genes are responsible for the specification and functions of each regulatory cell type, and that misexpression of these genes is capable of reprogramming regulatory cell types in accordance with this code.

In **Chapter 3,** I show that cells with AC potential are capable of reprogramming into every regulatory cell type, and that cells with LC potential differentiate instead into ACs in the absence of their endogenous bHLH proteins. In addition, I show that ectopic expression of HLH-12 alone

is sufficient to activate the endogenous migration mechanism of the LC and hDTC, and cause a cell to migrate while retaining its original fate.

In **Chapter 4,** I perform phylogenetic analyses to determine the extent of Class II bHLH gene conservation across various nematode species. I find that most Class II genes are highly conserved, with the exception of *hlh-12*.

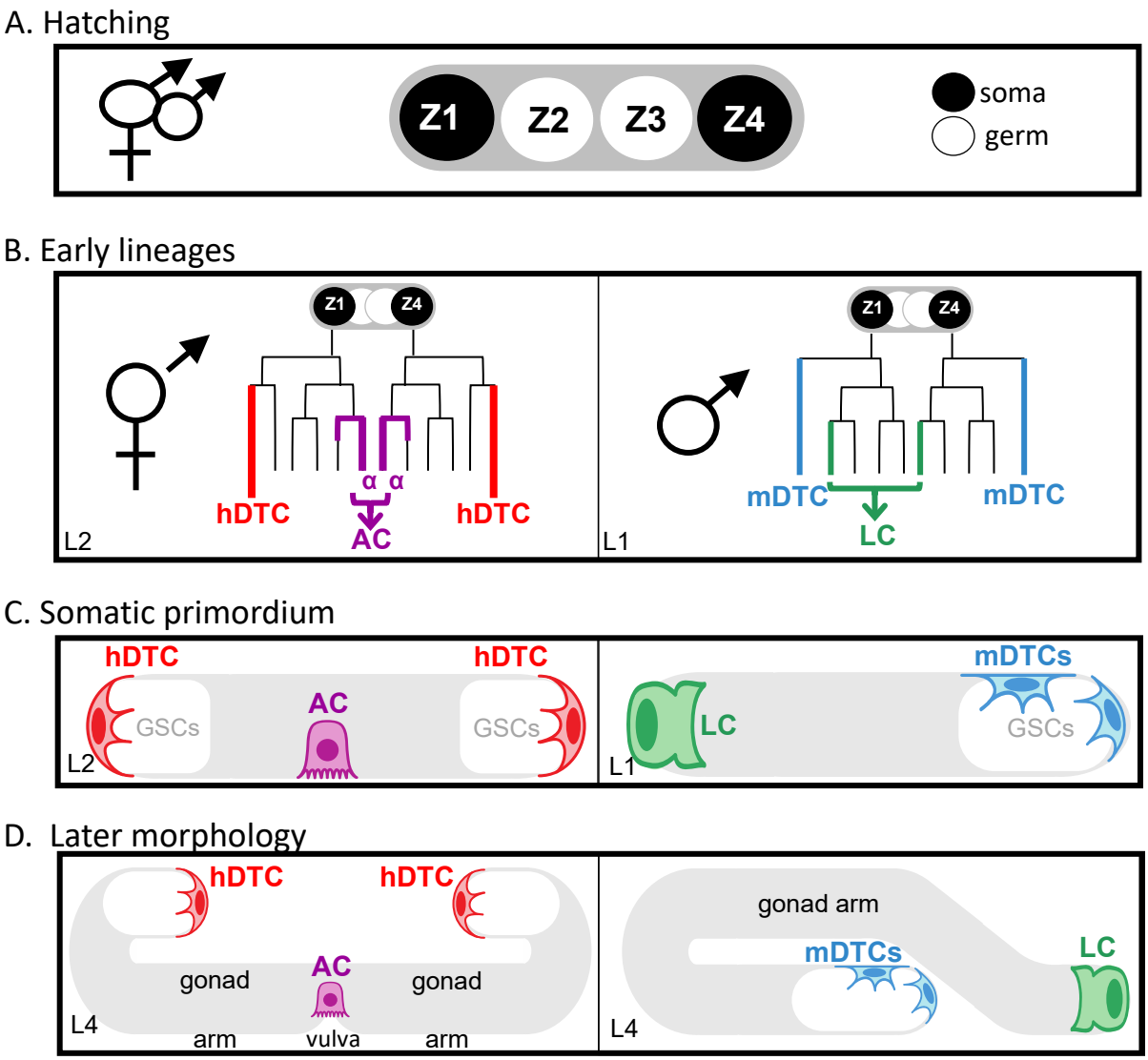


Figure 1.1. *C. elegans* gonadogenesis and somatic gonad primordium formation. Adapted from Sallee et al., 2017.

A. Males and hermaphrodites have the same gonad precursors at hatching. The somatic progenitors are Z1 and Z4; the germline progenitors are Z2 and Z3.

B. Lineages of somatic gonad cells up to the time of somatic primordium establishment.

C. Somatic primordium formation, with differentiated regulatory cells.

D. Later morphologies.

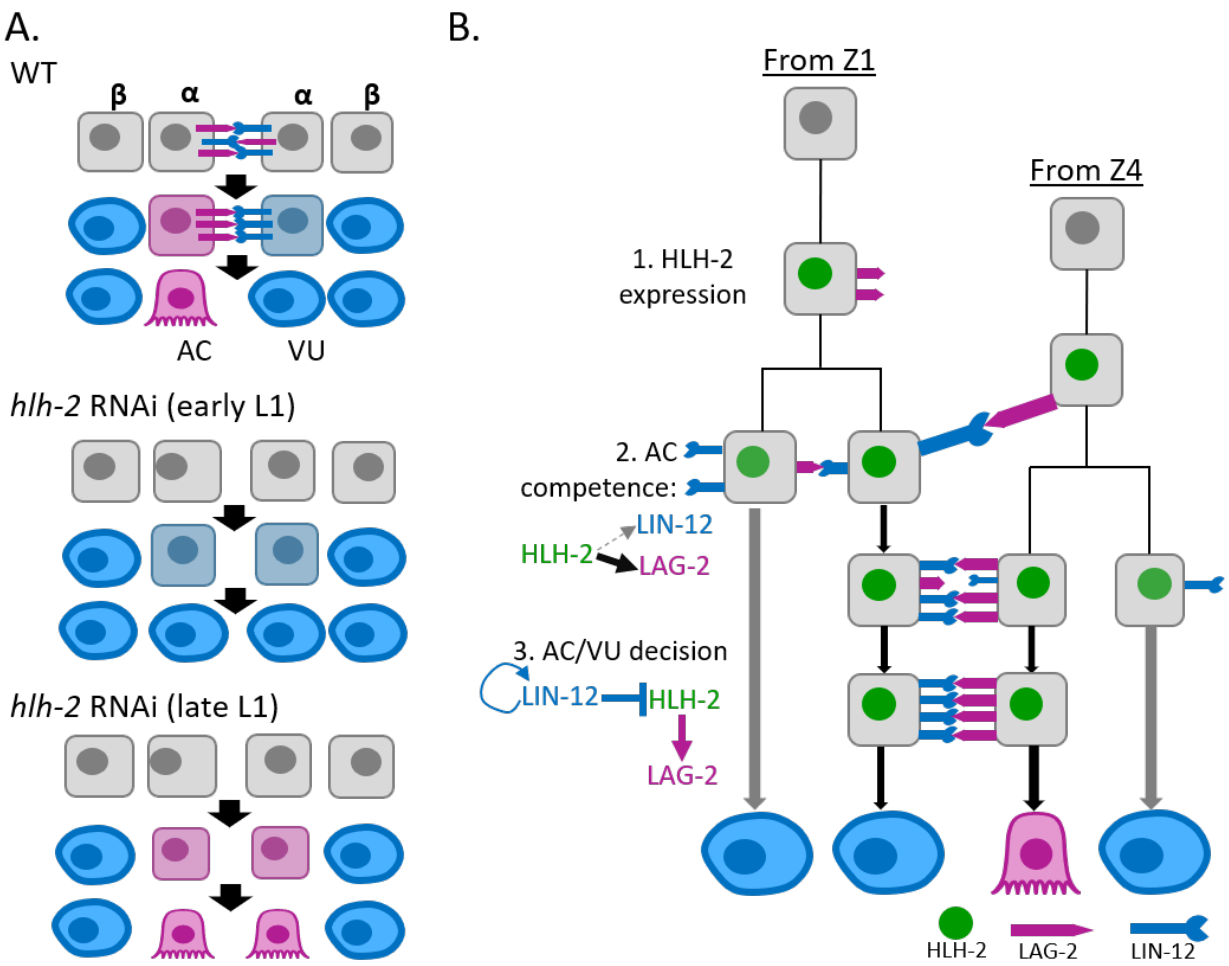


Figure 1.2. The AC/VU decision and AC specification. Adapted from Attner et al., 2019; Karp & Greenwald, 2004.

A. The wild-type AC/VU decision (top) and improper decisions after *hlh-2* RNAi performed in early L1 (middle) or late L1 (bottom). Genetic circuit shown only in top panel.

B. Birth order bias and *hlh-2* expression onset determine AC fate. In this schematic, Z1 is the first lineage to divide, and the Z4 lineage gives rise to the AC. In 2, dashed grey arrow: indirect activation; solid black arrow: direct activation.

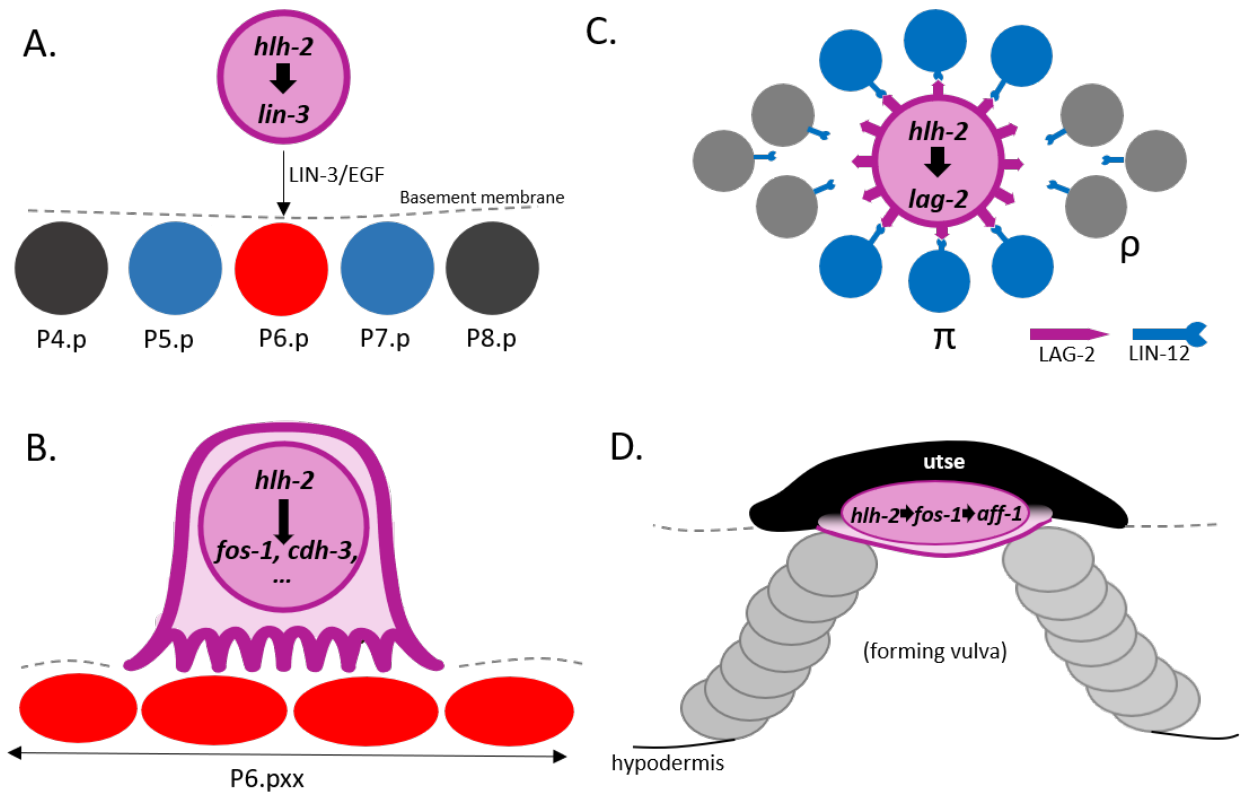


Figure 1.3. *hlh-2* controls aspects of AC function. The AC is represented in purple in each panel; in A and B, cells are shown as nuclei only. Adapted from Alper & Podbilewicz, 2008; Medwig-Kinney et al., 2020; Newman et al., 1995;Sallee & Greenwald, 2015; Sapir et al., 2007; Sternberg, 2005.

A. Induction of VPC fate by the AC. Red: primary fate; blue: secondary fate, grey: tertiary fate. Not shown is P3.p, which also has tertiary fate.

B. π cell fate decision (dorsal view). Blue cells are π cells; grey cells are ρ cells.

C. Invasion of the primary-fated P6.p descendants.

D. Fusion with the utse.

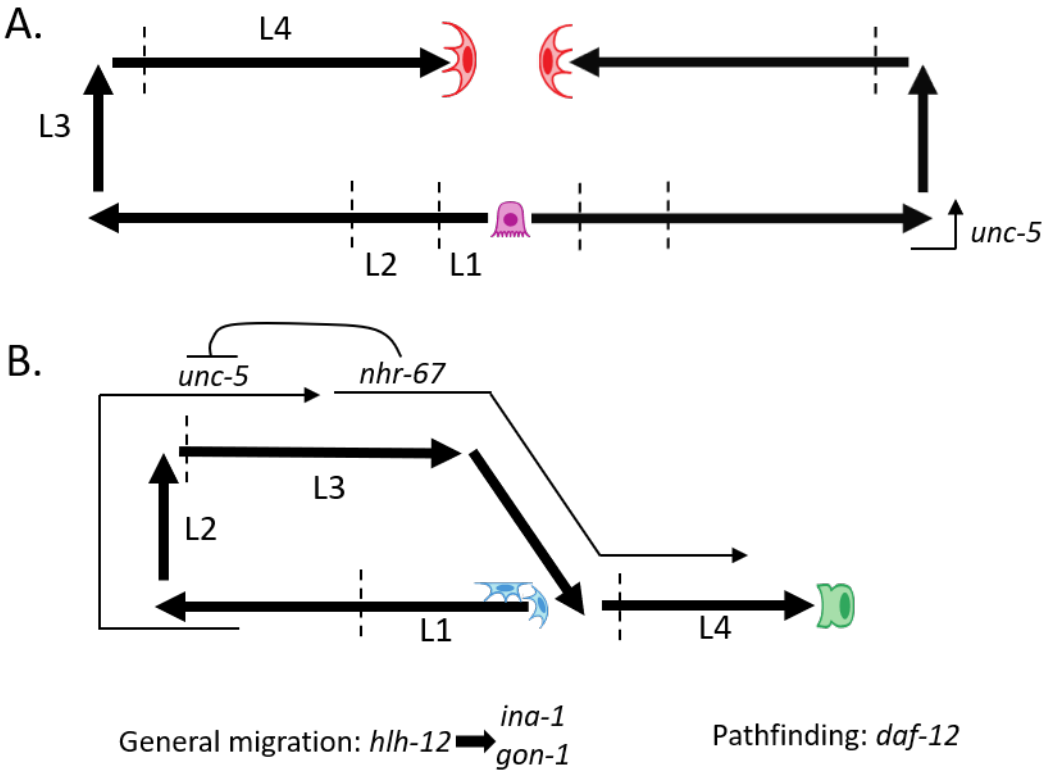


Figure 1.4. Migration of the hDTCs (A) and LC (B). Only selected genes regulating their migration are shown. General migration and pathfinding genes at bottom control migration of both the hDTCs and LC. Adapted from Antebi et al., 1998; Blelloch & Kimble, 1999; Kato & Sternberg, 2009; Meighan & Schwarzbauer, 2007; Wong & Schwarzbauer, 2012.

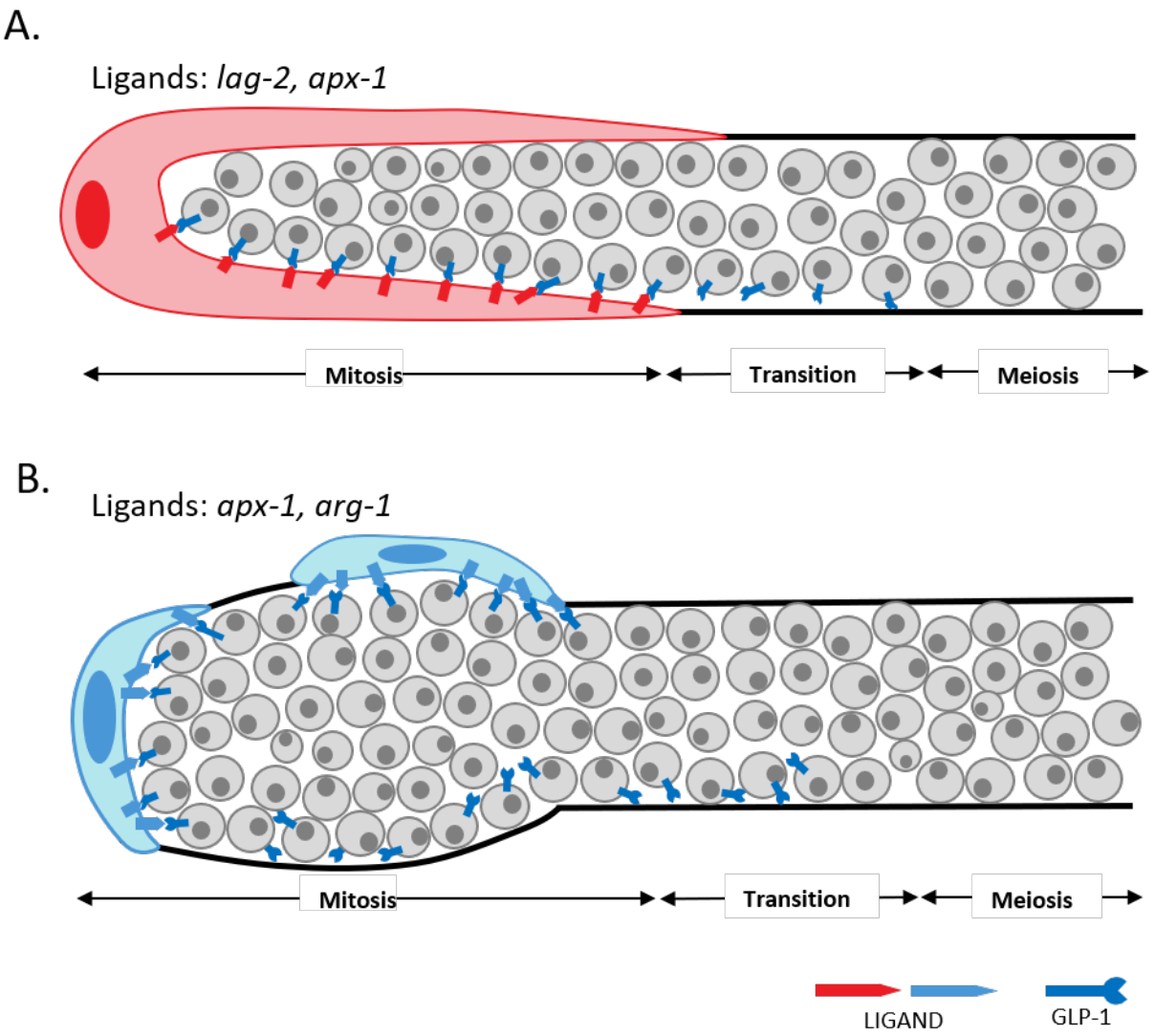


Figure 1.5. hDTC (A) and mDTC (B) niches. Diagrams represent cross-section of germline. Adapted from Crittenden et al., 2019.

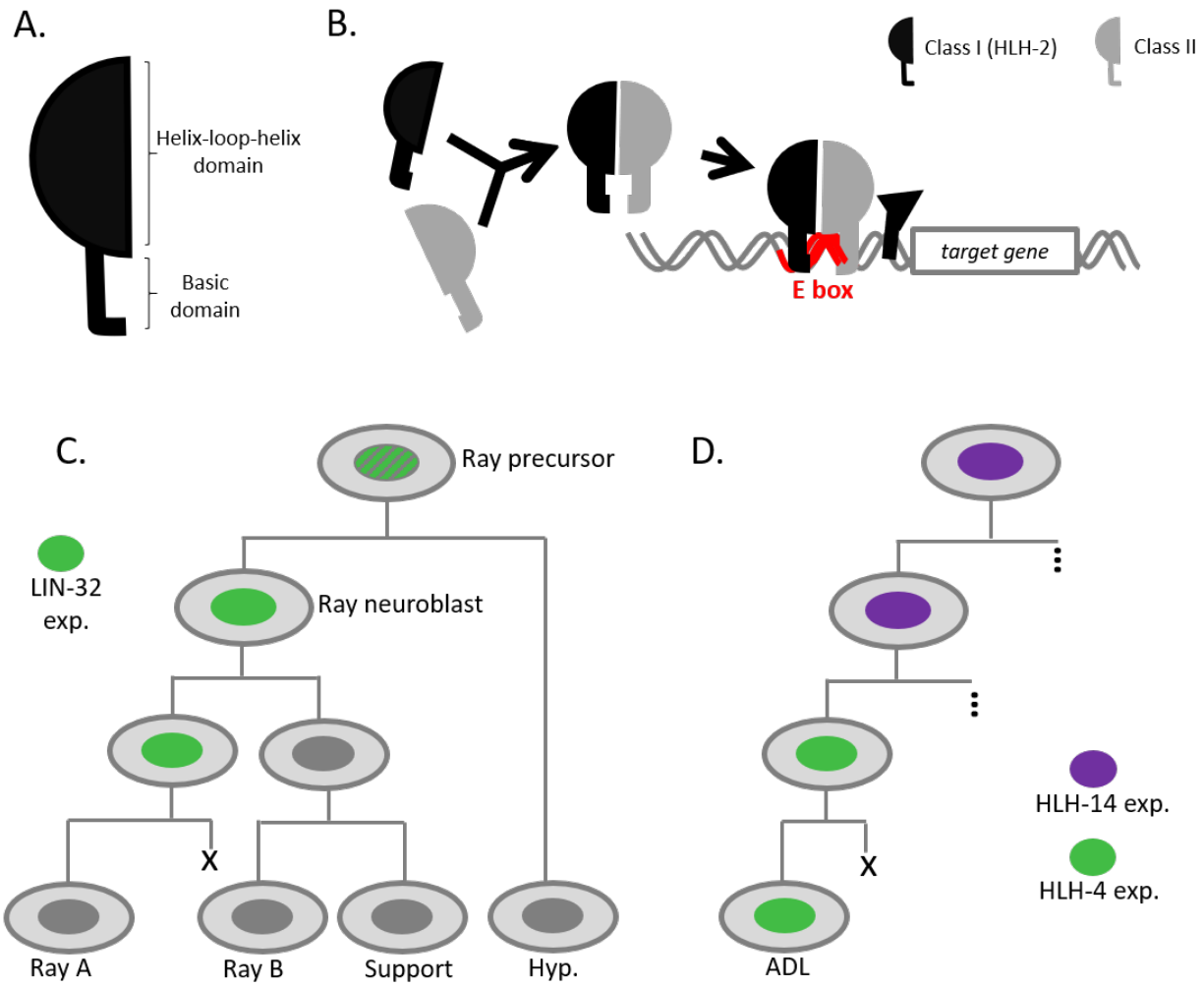


Figure 1.6. bHLH protein structure and function in cell fate specification. Adapted from Maria Sallee (personal communication); Masoudi et al., 2018; Portman & Emmons, 2000.

A. Cartoon of bHLH protein structure, which will be used throughout this thesis.

B. Class I monomers can dimerize with either other Class I or Class II monomers to promote transcription.

C. LIN-32 function in specifying male tail A-type ray neurons.

D. HLH-14 and HLH-4 function in specifying ADL. C, D: X reflects cell death.

Table 1.1. Markers and phenotypes used to assess specification and functions of regulatory cell types. **cdh-3* is involved in connection, but we used it to assess specification only. *** *hlh-2prox* was used to assess AC potential, not AC fate *per se.* Adapted from Sallee et al., 2017.

	Specification	Leader	Niche	Connect
AC	*lag-2, cdh-3*, hlh-2prox***			*fos-1, zmp-1*
hDTC	*lag-2, hlh-2, ceh-22b, hlh-12*	Outgrowth *ina-1*	*lag-2*	
LC	*lag-2, hlh-12*	Outgrowth		N/A
mDTC	*hlh-2*		*arg-1, apx-1*	

Chapter 2: A "bHLH code" for sexually dimorphic form and function of the *C. elegans* somatic gonad

My contributions to this work

Main text:

Figure 2.3B: bHLH genes are required for mDTC development

Figure 2.4:

A: Subtracting *hlh-3* converts the male-and-proximal LC to a female-and-proximal AC-like cell.

B: Addition of HLH-8+LIN-32 reprograms the female-and-proximal AC into a male-and-distal mDTC

(joint) C: Model for specification of regulatory cells of the somatic gonad primordium.

Supplemental text:

Figure S2.2B: *hlh-2(L2-RNAi)*: *hlh-2* is not required for *lag-2* expression in hDTCs after hDTC specification

Figure S2.3:

A: *lag-2* fosmid expression

B: *lag-2(7.2kb)p* expression

C: *naEx956[apx-1p]* expression

D: *ccIs4443[arg-1p]* expression

F: *lin-32(0)* mDTCs are specified

Figure S2.4: Gonad primordium formation in reprogramming experiments and compromised terminal AC function in LIN-32+HLH-8 “addition” hermaphrodites.

A "bHLH code" for sexually dimorphic form and function of the *C. elegans* somatic gonad

Maria Sallee[a,b]*, Hana Littleford[c]* and Iva Greenwald[b,c]†

[a]Dept. of Genetics and Development
[b]Dept. of Biochemistry and Molecular Biophysics
[c]Dept. of Biological Sciences
Columbia University
New York, New York 10027

*These authors contributed equally to this work
†Lead Author

Keywords: sexual dimorphism, cell fate specification, cell fate reprogramming

Summary:

Most animals have sexually dimorphic gonads; how such differences are generated is a fundamental question at the interface of developmental and evolutionary biology (Kopp, 2012; Zhu et al., 2014; Emmons, 2013). In *C. elegans*, sexual dimorphism in gonad form and function largely originates in different apportionment of roles to three "regulatory cells" of the somatic gonad primordium that forms in young larvae. Their essential roles include leading gonad arm outgrowth, serving as the germline niche, connecting to epithelial openings, and organizing reproductive organ development. The development and function of the regulatory cells in both sexes requires the basic Helix-Loop-Helix (bHLH) transcription factor HLH-2, the sole ortholog of the E proteins mammalian E2A and Drosophila Daughterless (Krause et al., 1997; Karp & Greenwald, 2003; Karp & Greenwald, 2004; Tamai & Nishiwaki, 2007; Chesney et al., 2009), yet how they adopt different fates to execute their different roles has been unknown. Here, we show that each regulatory cell expresses a distinct complement of bHLH-encoding genes--and therefore distinct HLH-2:bHLH dimers--and formulate a "bHLH code" hypothesis for regulatory cell identity. We support this hypothesis by showing that the bHLH gene complement is both necessary

and sufficient to confer particular regulatory cell fates. Strikingly, prospective regulatory cells can be directly reprogrammed into other regulatory cell types simply by loss or ectopic expression of bHLH genes, and male-to-female and female-to-male transformations indicate that the code is instructive for sexual dimorphism. The "bHLH code" appears to be embedded in a bow-tie regulatory architecture (Nelson et al., 2011; Friedlander et al., 2015), wherein sexual, positional, temporal, and lineage inputs connect through bHLH genes to diverse outputs for terminal features, and provide a plausible mechanism for the evolutionary plasticity of gonad form seen in nematodes (Felix & Sternberg, 1996; Kiontke et al., 2007; Matson & Zarkower, 2012; Zarkower, 2006; Sommer, 2005).

Results

Systematic analysis of potential heterodimerization partners for HLH-2 in gonadogenesis suggests the bHLH code hypothesis

The regulatory cells in hermaphrodites are the Anchor Cell (AC) and hermaphrodite Distal Tip Cells (hDTCs); in the male, they are the Linker Cell (LC) and the male Distal Tip Cells (mDTCs). Their lineal origins, the main roles investigated in this study, and marker gene expression and morphological features are depicted in Figure. 1.

At the outset of this study, we knew that *hlh-2* is critical for the development of all of the regulatory cells (Krause et al., 1997; Karp & Greenwald, 2003; Karp & Greenwald, 2004; Tamai & Nishiwaki, 2007; Chesney et al., 2009) and that HLH-2, like all E proteins, functions in a dimer (Massari & Murre, 2000; Grove et al., 2009). We also knew that HLH-2 acts as a homodimer for specification and function of the AC (Sallee & Greenwald, 2015); as E proteins more commonly form heterodimers with other bHLH proteins, we expected that heterodimers would mediate

various roles in other regulatory cells. Although HLH-12:HLH-2 heterodimers had been shown to mediate full outgrowth of gonad arms in both sexes (Tamai & Nishiwaki, 2007), no other bHLH partners had been implicated in regulatory cell specification or function. We therefore systematically assessed potential heterodimerization partners by functional and/or expression analyses (STAR Methods; Fig. S1, Table S1), and discovered that each regulatory cell type expresses an overlapping but distinct set of bHLH proteins soon after they are born and generally continuing throughout development—the same pattern as HLH-2, and described further below. Based on this unexpected observation, we hypothesized that the distinct complement of bHLH genes expressed in each regulatory cell constituted a "bHLH code" for regulatory cell fate (Fig. 1E). To test this hypothesis, we assessed if the bHLH gene complement is both necessary and sufficient to confer regulatory cell fate.

lin-32 and *hlh-12* are functionally redundant for hDTC development in accordance with the proposed code

In hermaphrodites, the hDTCs lead outgrowth of the two gonad arms underlying the sexually dimorphic U-shape, and serve as the germline niche by producing ligands that activate GLP-1/Notch in the germline stem cells (Kimble & Hirsh, 1979; Austin & Kimble; Henderson et al., 1994; Fitzgerald & Greenwald, 1995; McGovern et al., 2009). We assessed bHLH genes for roles in hDTC development using a sensitized, phenotype-based RNAi screen (for all predicted bHLH genes) and an expression screen of Class II bHLH genes, common heterodimerization partners for E proteins (Figs. S1, S2; Table S1; Massari & Murre, 2000; Sallee & Greenwald, 2015). Both approaches suggested that the code for the hDTC is LIN-32+HLH-12. LIN-32, like its orthologs mammalian ATOH1 and Drosophila Atonal, had been previously implicated in

nervous system development (Zhao & Emmons, 1995; Portman & Emmons, 2000; Miller & Portman, 2011); HLH-12 had been previously implicated in hDTC leader function in later larval stages (Tamai & Nishiwaki, 2007). Neither gene had been implicated in specification, early leader function, or niche function.

Fosmid-based reporters encoding GFP-tagged LIN-32 or HLH-12 suggested roles in specification and terminal functions: both genes are expressed in the hDTCs but not in their parents, and continue to be expressed in hDTCs throughout development, similar to both a fosmid-based GFP-HLH-2 reporter (Fig. S1) and endogenous HLH-2 (Karp & Greenwald, 2004). Although *lin-32(0)* and *hlh-12(0)* single mutants have two hDTCs, as in wild-type, we find that *hlh-12* and *lin-32* are redundantly required for hDTC specification: 63% (40/64) of prospective hDTCs of *hlh-12(0); lin-32(0)* double mutants exhibit both leader failure concomitant with failure to express markers of hDTC specification (Fig. 2A,B,D). In addition, *hlh-12* and *lin-32* are also functionally redundant for early leader function: even when hDTCs are present in *hlh-12(0); lin-32(0)* hermaphrodites based on marker expression, gonad arm outgrowth fails completely (n=23/64), a more extreme phenotype than the late outgrowth defect of *hlh-12(0)* and the normal outgrowth in *lin-32(0)* single mutants (Fig. 2B,D). Finally, *hlh-12* also promotes expression of the niche ligand *apx-1* (Fig. 2C).

In sum, LIN-32 and HLH-12, along with their obligate heterodimerization partner HLH-2, are required for hDTC specification and its terminal leader and niche functions, consistent with a LIN-32+HLH-12 code for hDTCs.

hlh-3, unique to the LC, confers its identity and blocks its inappropriate programming as an anchor cell in males

In males, the mDTCs provide germline niche function but do not have the leader role associated with the hDTCs. Instead, leader function is assigned to the LC (Fig. 1), which leads outgrowth of a single gonad arm to produce the sexually dimorphic J-shape, while the mDTCs fix the distal end in the midbody (Kimble & Hirsh, 1979). In its leader function, the LC resembles the hDTCs, but in forming the connection with epithelial cells, the LC parallels the AC. Also like the AC, the LC has a proximal lineage origin and is specified via *lin-12*/Notch signaling between bipotential precursors (Kimble & Hirsh, 1979; Greenwald et al., 1983).

The LC expresses three potential heterodimerization partners for HLH-2: HLH-12 and LIN-32, as in the hDTCs, but also HLH-3 (of the Drosophila Achaete-Scute/mammalian ASCL family). In *hlh-12(0)* males, the LC is present and begins to lead gonad arm outgrowth, but often does not execute its posterior turn (17/40 males), similar to the hDTC outgrowth defect observed in *hlh-12(0)* hDTCs (Tamai & Nishiwaki, 2007). In contrast, loss of *hlh-3*, the only bHLH gene uniquely expressed in the LC and not in any other regulatory cells, causes profound leader failure in 100% (40/40) of males, distinctly more severe than the leader outgrowth abnormality of *hlh-12(0)* males (Fig. 3A). We examined the somatic primordium for evidence a LC is present by looking at *lag-2*, which in wild-type males, is expressed in the LC but not mDTCs (Fig. S3A, S3B). We observed a single cell in *hlh-3(0)* mutants that expresses *lag-2* (100%, n>20), consistent with a specified LC that lacks leader function. However, the *lag-2*-expressing cell generally remains together with the vas deferens precursor cells instead of taking the typical LC position at one end of the somatic primordium. We therefore wondered if loss of *hlh-3* resulted in *lag-2* expression in

the female mode, as in the AC or hDTCs (Fig. S3A, S3B), or further, if "subtraction" of *hlh-3* reprograms the LC into an AC-like cell.

Our analysis suggests that the "unmigrated" cell in *hlh-3(0)* males has AC-like properties but is not completely transformed into an AC. To test if the unmigrated cell in *hlh-3(0)* males has AC features, we first examined *hlh-2*prox::*gfp*. In hermaphrodites, this marker, a fragment of the *hlh-2* promoter, is expressed in the four proximal cells of the somatic gonad primordium that have the potential to be the AC (the "α" cells and their sisters; Fig. 1D, Fig S4B), and remains strongly expressed in the committed AC; it is not expressed in the committed LC or cells with LC potential, hDTCs, or mDTCs (Sallee & Greenwald, 2015; Fig. 4A). In *hlh-3(0)* males, *hlh-2*prox::*gfp* is expressed solely in the single *lag-2*-expressing unmigrated cell, suggesting that "subtraction" of *hlh-3* reprograms the prospective LC into an AC-like cell (n=9/20, Fig. 4A). This transformation is partial, however, as the unmigrated cell does not express a later marker of AC differentiation, *cdh-3::gfp* (data not shown).

Why does subtraction of *hlh-3* convert the LC into an AC-like cell (code: HLH-2 only) instead of an hDTC-like cell (code: LIN-32+HLH-12)? A simple hypothesis is that *hlh-3* also cross-regulates expression of one or both of the other bHLH genes. We could not examine *lin-32* because the characterized *lin-32* fosmid reporter *otIs594* was inviable in combination with *hlh-3(0)*, but remarkably, the HLH-12 fosmid-based translational reporter is not expressed in the *hlh-3(0)* prospective LC (n=16/18 early L2 males), indicating that in actuality, the unmigrated cell does not have the hDTC code.

Ectopic expression of the mDTC code, LIN-32+HLH-8, reprograms presumptive anchor cells into mDTCs

The mDTCs express LIN-32 and HLH-8, the ortholog of Drosophila and mammalian Twist (Grove et al., 2009). Although the mDTCs, like hDTCs, provide niche function (Kimble & White, 1981), our analysis suggests that the niche cells display sexual dimorphism in expression of ligands for GLP-1/Notch; notably, a transcriptional reporter for *arg-1* (Kostas & Fire, 2002), a Notch ligand gene, is expressed in the mDTCs and not in the hDTCs or any other regulatory somatic gonad cells, allowing us to distinguish the mDTCs from the hDTCs (Fig. S3D). Using a *lin-32* null allele and an *hlh-2* temperature-sensitive allele, we found that bHLH genes promote expression of *arg-1* (Fig. 3B). We could not assess the contribution of *hlh-8* using loss-of-function analysis because *hlh-8(0)* males do not live much beyond the L2 stage combined with our markers, and the mDTCs are refractory to RNAi even in sensitized backgrounds (Fig. S3E). Instead, we performed an "addition" experiment, as described next, that revealed that *hlh-8* and *lin-32* together specify mDTC identity.

The sexual identity, position, lineage history, and developmental stage of the regulatory cells has been established before *hlh-2* is expressed, and its activity specifies their fates (Karp & Greenwald, 200; Chesney et al., 2009; Chang et al., 2004; Lam et al., 2006). The timing and specificity of reporters for *hlh-2* and the bHLH genes encoding its dimerization partners suggest that their regulatory regions integrate these inputs. Indeed, a specific element of the *hlh-2* promoter region, *hlh-2*prox, is necessary and sufficient for its expression in the AC and its precursors in a wild-type hermaphrodite (Sallee & Greenwald, 2015; Fig. 4A). Thus, a corollary of the code hypothesis is that bypassing these inputs by ectopically expressing the bHLH gene complement of

one regulatory cell type should be sufficient to reprogram that cell into the regulatory cell type normally specified by that gene complement.

The AC provides a *tabula rasa* for such an "addition experiment" because HLH-2 functions as a homodimer to specify its fate and functions, and we could not identify any potential dimerization partner in a functional assay or by Class II bHLH gene expression (Fig. S1). We therefore tested if we could switch the identity of the AC, which has a proximal lineage origin and a female sexual identity, to the mDTC, which has a distal lineage origin and a male sexual identity. To do so, we used *hlh-2*prox to drive expression of HLH-8 and LIN-32 individually or together in the AC (Fig. 4B). Ectopic expression of HLH-8, but not LIN-32, is sufficient to induce strong expression of the mDTC marker *arg-1*p::*gfp* in the AC; however, in both cases, AC morphology, and later uterine and vulval morphology, appear normal, indicating that AC specification and function is not affected (Fig. S4).

Remarkably, combining the transgenes that express HLH-8 and LIN-32 causes proximal gonadal cells not only to express the *arg-1*p::*gfp* mDTC marker, but also to extend processes towards the germ line, characteristic of DTCs but not of ACs (Fig. 4B). We see as many as four adjacent proximal cells with this morphology, consistent with the reprogramming of all four cells that have the potential to be ACs into mDTC-like cells. In addition, the hermaphrodites display developmental defects consistent with defects in AC specification or terminal fate: complete failure of vulval induction, delayed vulval induction, abnormal vulval morphogenesis, and defective uterine seam (utse) formation (Fig. S4C-G).Together, these results indicate that cells with AC potential are directly reprogrammed into mDTCs in the presence of HLH-8 and LIN-32.

Discussion

In sum, the analysis presented here indicates that sexual dimorphism of the regulatory cells of the *C. elegans* somatic gonad reflects a "bHLH code" in which a distinct set of bHLH transcription factors specify each regulatory cell type and/or distinctive terminal features associated with it.

The expression pattern and pervasive functions of the bHLH genes that constitute the code suggest that they reside in classic “bow-tie” regulatory circuits, wherein multiple inputs and outputs connect through a relatively simple intermediate (Nelson et al., 2011; Friedlander et al., 2015). The different inputs include sexual, positional, temporal, and lineage information that leads to the regulatory cell-specific expression of particular bHLH genes (Fig. 4C). The ability to bypass these inputs to directly reprogram the proximal-female AC into the distal-male mDTC—the “addition” experiment of Fig. 4—supports this inference. In addition, since all four cells with AC potential appear to be transformed into mDTCs, this observation further suggests an essential equivalence of the prospective regulatory cells before their fates are specified by the expression of their specific bHLH complement.

In turn, different combinations of bHLH genes in each type of regulatory cell direct the different outputs—bHLH transcriptional target genes. It is known that HLH-2 homodimers directly regulate many different genes in the anchor cell (Karp & Greenwald, 2003; Sallee & Greenwald, 2015; Hwang & Sternberg, 2004; Hwang et al., 2007; Verghese et al., 2011). It is intriguing that HLH-12 is expressed in hDTCs and the LC, which both have leader function, and at least two genes required for outgrowth appear to be bHLH transcriptional targets (Tamai & Nishiwaki, 2007; Meighan et al., 2015). Other outputs may be more complex, and reflect differences in bHLH heterodimer binding preferences (Grove et al., 2009), or incorporate other

information, such as feed-forward influences from the input pathways on target gene availability or selection, and cross-regulation of bHLH genes, as observed in the LC (described above).

The bHLH code also has implications for evolution of gonad form and function. The gonad primordium of all nematodes is remarkably similar, yet, in addition to sexual dimorphism within a species, there is striking plasticity in the lineages that give rise to the regulatory cells and in the distribution of their roles when different species are compared (Felix & Sternberg, 1996; Sommer & Sternberg, 1994). For example, in *Mesorhabditis*, the AC does not induce the vulva and instead leads outgrowth of the single gonad arm posteriorly towards the vulva (Felix & Sterberg, 1996), reminiscent of the *C. elegans* LC. The bHLH code we have deduced in *C. elegans* provides a plausible mechanism for how the key roles of specialized cells may be grouped as a genetic module and reassigned to other cells in evolution. Indeed, the bHLH code is reminiscent of the Hox code—the canonical example of a group of related genes that control morphological diversity in organisms and are modulated during evolution, with "executive functions" at higher levels, including cross-regulation amongst themselves, as well as directly regulating "blue collar" or "realisator" genes that execute terminal functions (Pearson et al., 2005; Garcia-Bellido et al., 1973).

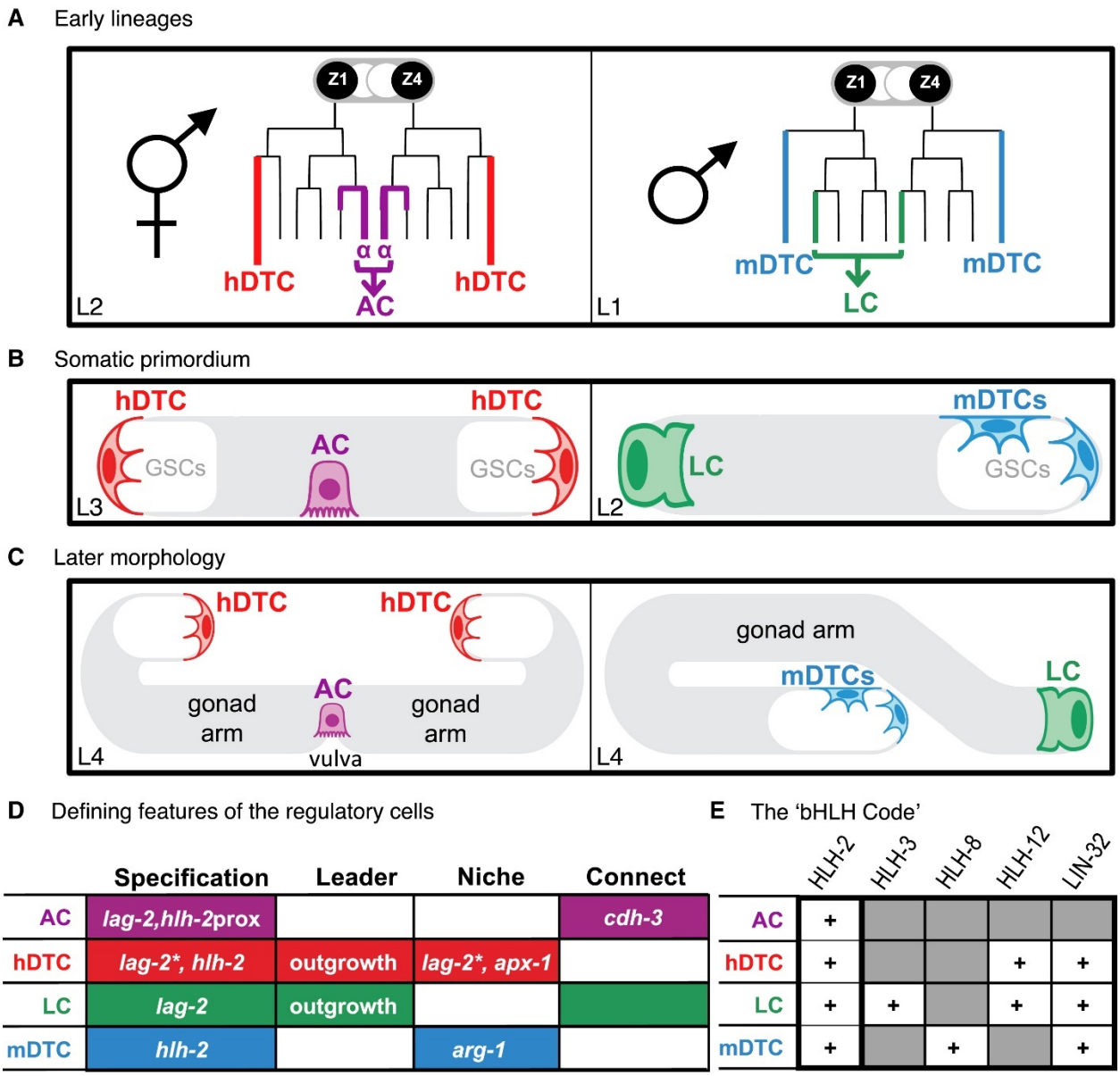
Finally, we note that bHLH proteins are highly conserved in structure and function—indeed, the human ortholog of HLH-2 can substitute for HLH-2 in *C. elegans* proximal gonadogenesis (Sallee & Greenwald, 2015)—and while there is great variation in sex determination mechanisms, there is deep conservation of the Doublesex/Mab-3 regulators (Gamble & Zarkower, 2012). Thus, the fundamental principles that underlie gonadogenesis in nematodes may be relevant to the evolution of sexually dimorphic form and function in all animals.

Author contributions

M.D.S. and H.L. conducted the experiments; M.D.S., H.L. and I.G. analyzed and interpreted the data, and prepared the manuscript.

Acknowledgements

We are grateful to Oliver Hobert, Neda Masoudi, Douglas Portman, and Marian Walhout for reagents; Taner Aydin and Davys Lopez for assistance; Oliver Hobert for discussion; and Claude Desplan, Michelle Attner, Claire de la Cova, Xantha Karp, and David Matus for critical reading of the manuscript. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). This study was supported by NIH grant R01 GM115718 (to I.G.).



	Specification	Leader	Niche	Connect
AC	*lag-2,hlh-2prox*			*cdh-3*
hDTC	*lag-2*, hlh-2*	outgrowth	*lag-2*, apx-1*	
LC	*lag-2*	outgrowth		
mDTC	*hlh-2*		*arg-1*	

	HLH-2	HLH-3	HLH-8	HLH-12	LIN-32
AC	+				
hDTC	+			+	+
LC	+	+		+	+
mDTC	+		+		+

Figure. 2.1. Gonadogenesis in *C. elegans*. See also Fig. S1.

A-C. Schematic representation of gonadogenesis.
A**.** The somatic precursors Z1 and Z4 generate the somatic primordium by the lineages shown: hDTC (red) and AC (purple) in hermaphrodite, and mDTC (blue) and LC (green) in male.

B. The regulatory cells are terminally differentiated and adopt characteristic positions; the remaining cells, located in the region in gray, are progenitors to structures that develop later. In hermaphrodites, the hDTCs provide the "niche function" of nurturing germ line stem cells (GSCs) and lead outgrowth of the gonad arms, first migrating outward in opposite directions to extend the gonad arms ("leader function"), and then reflexing to generate a U-shape. The two "α" cells

interact via LIN-12/Notch to resolve which will be the AC, which adopts a central position in the somatic primordium and remains there to induce the vulva, pattern the ventral uterus, and forge the connection between them. In males, the anterior prospective mDTC migrates posteriorly to join the other prior to primordium formation to establish the distal end and to serve as the niche for the GSCs; LIN-12/Notch also resolves which of two cells becomes the LC, which leads migration of the single arm outward and forges the connection with the cloaca.

C. By L4, the hDTCs and LC have led outgrowth of the gonad arms, and the AC has begun to induce vulval morphogenesis; the hDTCs and mDTCs continue to stimulate GSCs.

D. Gene expression profiles and cell functions of each regulatory cell. Our functional analysis was based on the terminal features or marker expression shown here. *bhlh* mutants frequently had LC migration defects, impeding an analysis of the LC "Connect" function. For details on transcriptional reporters, see the Key Resources Table and STAR Methods.

E. The "bHLH code" for sexually dimorphic gonad development. For details on bHLH reporters, see STAR Methods, Key Resources Table, and Fig. S1.

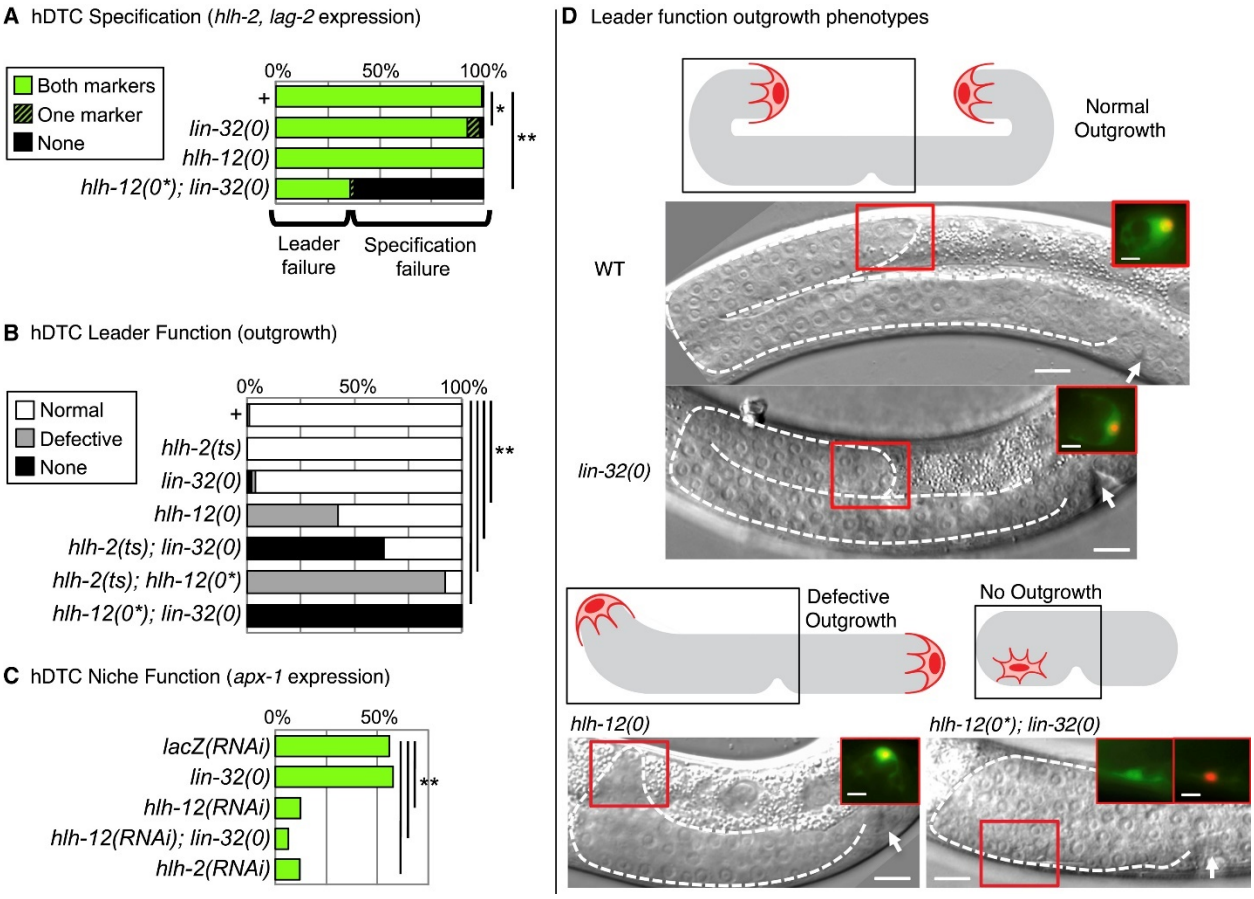


Figure 2.2. bHLH genes are required for hDTC specification and its terminal features. See also Fig. S2. Each hDTC was scored independently. "*hlh-12(0*)*" denotes that null homozygotes were obtained as segregants from balanced heterozygotes. See STAR Methods for additional details. Statistical analysis: two-tailed 2x3 Fisher's exact probability test, *p<0.01, **p<0.001.

A. Analysis of null mutants indicates that *hlh-12* and *lin-32* are required redundantly to specify hDTC fate. Percent of hDTCs expressing transcriptional reporters for *lag-2* and *hlh-2*. Normally, hDTCs express both *hlh-2* and *lag-2* transcriptional reporters (*arEx2194[hlh-2*p::*gfp]* and *arIs222[lag-2*p::*tagrfp])*. *lag-2* expression defines hDTC specification because it does not require *hlh-2* activity after specification (Fig. S2, STAR Methods). In *hlh-12(0); lin-32(0)* double mutants, a prospective hDTC either expresses both *hlh-2* and *lag-2* reporters (indicating specification) or neither reporter (indicating lack of specification).

B. *hlh-12* and *lin-32* are required redundantly for hDTC leader function. Normal, defective, or no outgrowth of the hDTC was assessed by examining gonad arm shape and hDTC position (criteria in D).

C. *hlh-2* and *hlh-12* are required for expression of a transcriptional reporter for the hDTC niche ligand *apx-1*, but *lin-32* is not. Percent of hDTCs expressing the transcriptional reporter *apx-1*p*::gfp* (*naEx156* (McGovern et al., 2009)). *hlh-2* and *hlh-12* RNAi leads to loss of *apx-1*p::*gfp*

expression in hDTCs, while the *lin-32(0)* loss does not. See Figs. S2 and S3 for further assessment of niche function.

D. Representative phenotypes scored in B. L4 hermaphrodite gonad morphology is shown. Top, normal hDTC function in wild-type or *lin-32(0)* hermaphrodites; bottom left, defective outgrowth, characterized by deficient elongation and/or turning, as in *hlh-12(0)* hermaphrodites (Tamai and Nishiwaki, 2007); bottom right, leader failure (no outgrowth), as in *hlh-12(0*); lin-32(0)* hermaphrodites. Photomicrographs show the anterior half of an L4 hermaphrodite gonad, outlined in white dotted lines; arrow, vulva. Red boxes indicate the position of the hDTC; the inset shows expression of two markers of hDTC specification: *hlh-2*p::*gfp* and *lag-2*p::*tagrfp*. Cytoplasmic GFP shows the presence of processes characteristic of hDTCs. In the *hlh-12(0*); lin-32(0)* hermaphrodite, GFP expression is often dim, and brightness and contrast were adjusted (left inset) to visualize hDTC processes. Scale bar is 10 μm, inset scale bar is 5 μm.

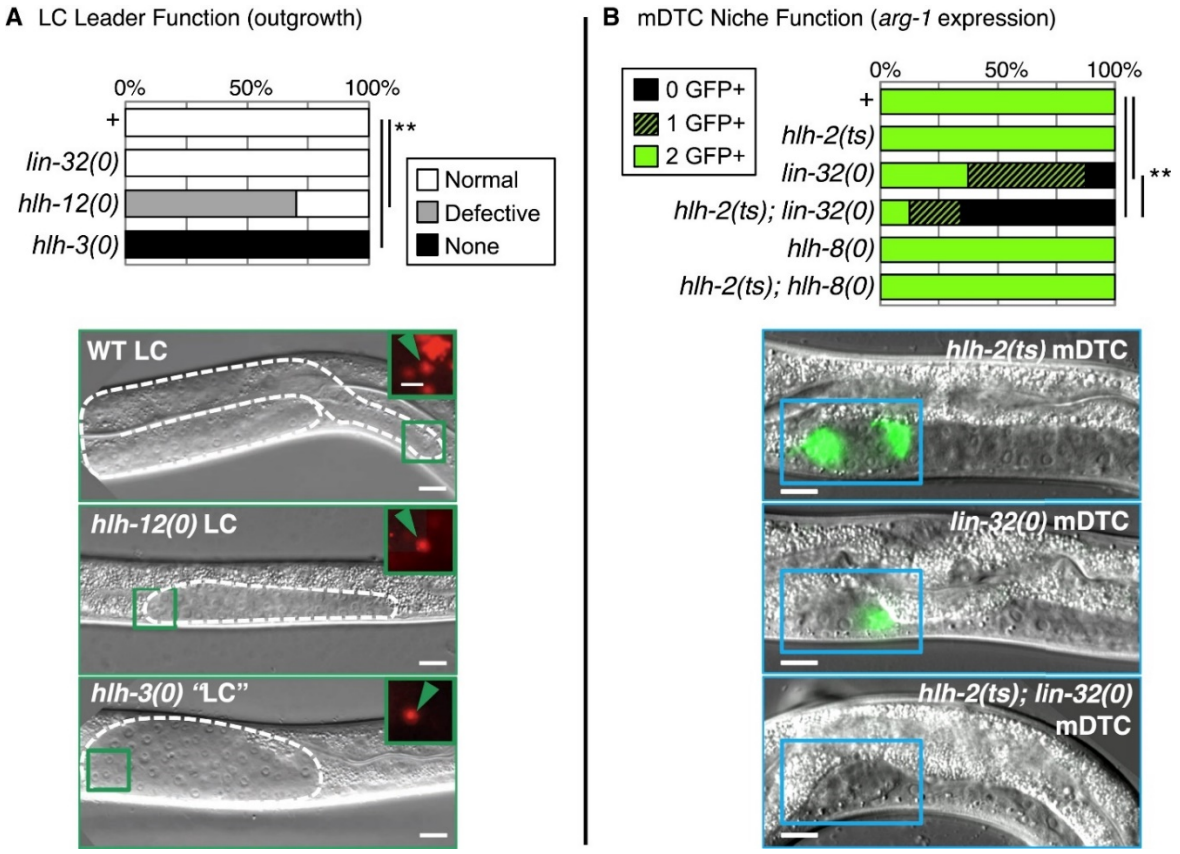


Figure 2.3. bHLH genes are required in males for terminal functions of the LC and mDTCs. See also Fig. S3. Statistical analysis in graphs: two-tailed 2x3 Fisher's exact probability test, *p<0.01, **p<0.001. For photomicrographs, scale bar is 10 μm, inset scale bar is 5 μm.

A. bHLH genes are required for LC development. Top, the percentage of bHLH mutant males with abnormal outgrowth, as in *hlh-12(0)* (Tamai and Nishiwaki; 2007), or no outgrowth, as in *hlh-3(0)*. In *hlh-3(0)* males, *lag-2* reporters are expressed, but the apparent failure of gonad arm outgrowth results from "LC"→"AC" transformation (Fig. 4), suggesting that *hlh-3* specifies LC fate. Bottom, representative photomicrographs show normal outgrowth, defective outgrowth, and leader failure. White dotted lines outline the gonad, and the green box and arrowhead indicate the position of the single *arIs222[lag-2*p::*tagrfp]*-expressing cell in the region of the somatic gonad shown in the inset.

B. bHLH genes are required for mDTC development. Top, the percentage of bHLH mutant males expressing *arg-1*p::*gfp* (*ccIs4443*). Two HLH-2-YFP-expressing cells are present in *lin-32(0)* males, suggesting mDTC fate is still specified (Fig. S3F). Strains with *lin-32(0)* were scored as L4-stage larvae; strains with *hlh-8(0)* were scored as L2 larvae, because males arrest around that time. The enhancement observed in *hlh-2(ts); lin-32(0)* males suggests that another bHLH gene is functionally redundant with *lin-32* for either specification or niche function. Fig. 4 shows that *lin-32* and *hlh-8* together cause AC→mDTC reprogramming, suggesting that they normally act together to specify the mDTC. Bottom, representative photomicrographs shows *arg-1* reporter expression, with blue boxes indicating the distal region of the L4 male gonad.

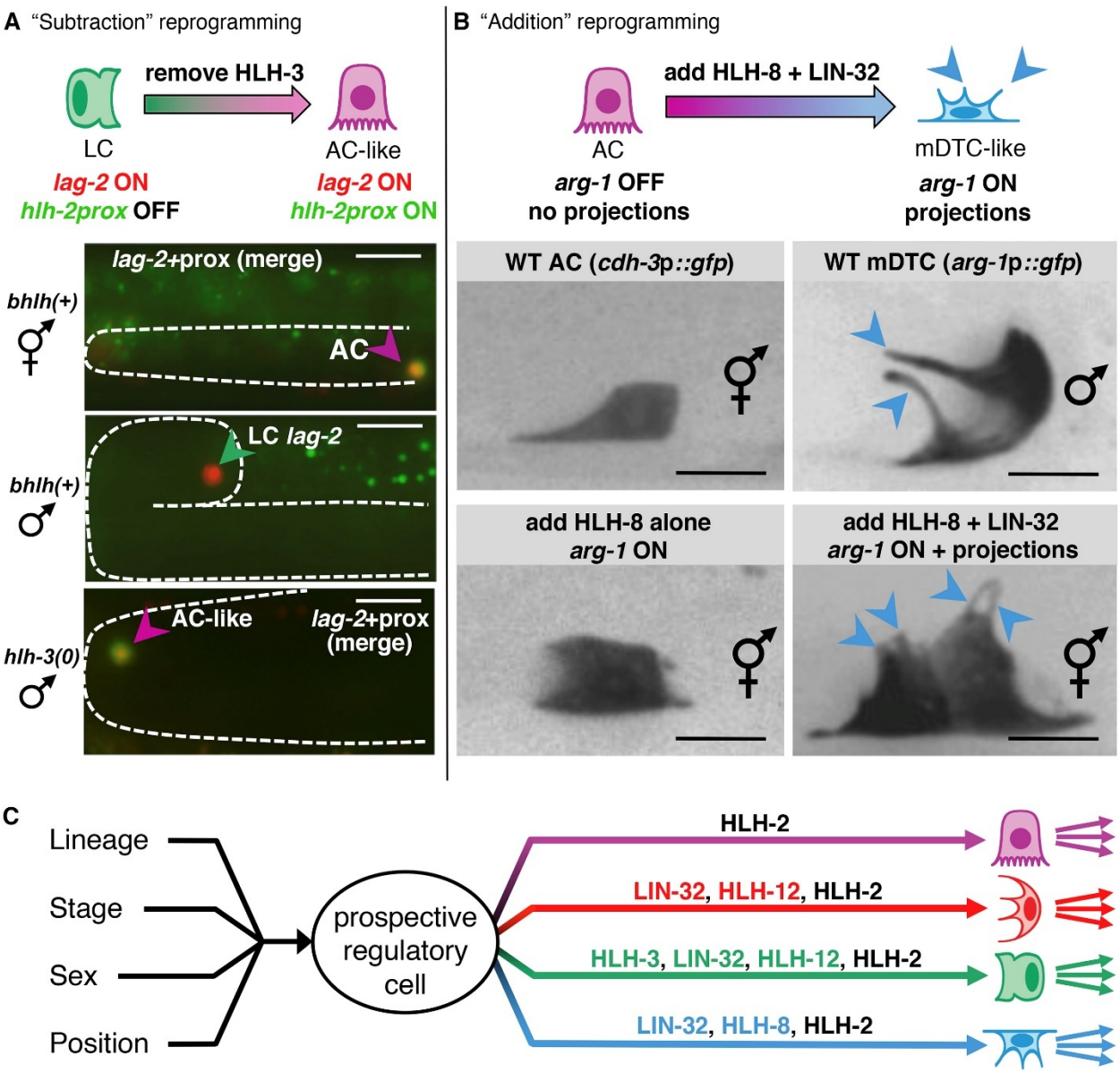


Figure 2.4. Reprogramming regulatory cell fate by subtraction or addition of bHLH gene activity in accordance with the "bHLH code." See also Fig. S4.

A. Subtracting *hlh-3* converts the male-and-proximal LC to a female-and-proximal AC-like cell. In wild-type, the AC expresses both *hlh*-2prox::*gfp* (green) and *lag-2::tagrfp* (red; purple arrowhead) and the LC expresses *lag-2::tagrfp* but not *hlh*-2prox::*gfp* (green arrowhead). In *hlh-3(0)* males, we observe a partial LC-to-AC conversion: *hlh-2*prox::*gfp* is ectopically expressed in the single *lag-2*-expressing cell (purple arrowhead, AC-like; n = 16/21).A dotted white line outlines the gonad, and scale bar is 10 μm.

B. Addition of HLH-8+LIN-32 reprograms the female-and-proximal AC into a male-and-distal mDTC. Normally, an AC (marked by *arIs51[cdh-3*p::*gfp]* expression) does not extend cellular processes toward the germ line (upper left), while a mDTC (marked with *arg-1*p::*gfp*) does (blue

arrowhead, upper right). Ectopic expression of HLH-8 in the presumptive AC causes only ectopic expression of *arg-1*p*::gfp* (lower left; n = 19/20), and LIN-32 alone has no overt effect (data not shown). However, when both HLH-8 and LIN-32 are expressed, we see 2-4 mDTC-like cells in the L2 stage, with both *arg-1*p*::gfp* expression (n = 20/20) and processes extending into the proximal germ line during the L2 stage (blue arrowheads, n = 13/15). In the individual shown (lower right), all four cells were transformed; the image shows a Z-projection of two of the four cells that express *arg-1*p*::gfp* and display mDTC-like germline processes. Interestingly, mDTC features are most apparent in the L2 stage, and appear to diminish with developmental progression—e.g., the number and extent of the mDTC-like processes decreases in later stages—suggesting that *hlh-2*prox expression may be lost when the cells are mDTC-like, concordant with the normal lack of *hlh-2*prox in such cells. Scale bar is 5 μm.

C. Model for specification of regulatory cells of the somatic gonad primordium.

Experimental model and subject details

C. elegans strains

See Key Resources Table for the full list of strains. Strains were maintained under standard conditions at 15°C, 20°C, or 22°C and scored at different temperatures as indicated. *pha-1(e2123ts)* facilitated the generation of some transgenes, and *him-8(e1489) IV* or *him-5(e1490)* V were included in some strains to generate self-progeny males. Hermaphrodite and male larvae and adults were scored, and sex and developmental stage are stated for each experiment. The RNAi sensitizer *nre-1(hd20) lin-15b(hd126)* (Schmitz et al., 2007) was included in most strains in which RNAi was performed. *arIs51[cdh-3*p::*gfp] IV* (Karp & Greenwald, 2003) and *oxTi414[eft-3*p::*mcherry] IV* (Frokjaer-Jensen et al., 2014) facilitated manipulations with *hlh-12*.

The following *bHLH* mutations were used: *hlh-2(bx108ts)* is a temperature-sensitive missense mutation that causes a low-penetrance 2AC defect at 25°C, and can genetically sensitize worms to the loss of other bHLH genes (Sallee & Greenwald, 2015; Portman & Emmons, 2000). *hlh-3(tm1688)* (Doonan et al., 2008)*, hlh-8(nr2061)* (Meyers & Corsi, 2010)*, hlh-12(tk68)* (Tamai & Nishiwaki, 2007), and *lin-32(tm1446)* (Miller & Portman, 2011) are deletion alleles that remove a large portion of the bHLH domain, making them likely null alleles.

The following transgenes were used as cell fate markers. See also Fig. S3 for further observations on the expression of *lag-2* and *arg-1* reporters.

We used three different *lag-2* reporters, depending on the color and specific features desired. All are strongly expressed in the AC, hDTCs, and LC (Fig S3). (i) *arIs222[lag-2*p::*2xnls-tagrfp]* (Sallee & Greenwald, 2015), often referred to here as *lag-2*p::*tagrfp*, contains 7.2 kb of 5' flanking region. (ii) *arIs131[lag-2*p::*2xnls-yfp]* (Li & Greenwald, 2010) also contains 7.2 kb of 5' flanking region. (iii) *arEx2499* is a complex array containing a *lag-2-gfp* fosmid translational

reporter, constructed as described below.

*ccIs4443[arg-1*p::*gfp]* is expressed in the mDTC (Kimble & White, 1981) but not in the other regulatory gonadal cells (Fig. S3D).

arIs51[*cdh-3*p::*gfp*] (Karp & Greenwald, 2003) is expressed in the AC, but not in the hDTC, mDTC, or LC.

naEx156[*apx-1*p::*gfp*] (McGovern et al., 2009) was used as a readout of hDTC niche function (Fig. 2). Due to a high degree of mosaicism, it was not used as a readout for mDTC niche function (Fig. S3C).

Like other *hlh-2*prox-based reporters (Sallee & Greenwald, 2015), *arTi22*[*hlh-2*prox::*gfp-his-58*] (this study) is expressed in the four cells with AC potential (the α and β cells) and their parents; when the somatic primordium is forming, expression persists more strongly in the two "α" cells, which also retain AC potential until it is resolved by LIN-12/Notch signaling; and then after the somatic primordium has formed, it becomes strongly expressed in the AC and weakly expressed in the VUs, but not in other gonadal cells (Fig. 4A).

arEx2194[hlh-2(5.2kb)p::*gfp*] (this study) is expressed in the hDTC and was used in conjunction with *arIs222* to indicate a specified hDTC (see Fig. 2).

The following transgenes were made to examine the consequences of ectopic expression of bHLH genes in the AC via *hlh-2*prox, shown in Fig. 4 and Fig. S4: *arTi148*, derived from pHL3, expresses LIN-32; and *arEx2500*, a complex array derived from pHL5, expresses HLH-8. The constructs used to make these transgenes are described in the next section.

Method details

Ectopic bHLH expression constructs

"*hlh-2*prox" is a fragment from the *hlh-2* 5' flanking region that, when the somatic

primordium is forming or formed, drives expression in the four proximal cells of hermaphrodites with AC potential, their parents, and the specified AC and VUs, with expression subsequently maintained relatively high in the AC and low in the VUs (Sallee & Greenwald, 2015; see also Fig. 1).

Using standard restriction site cloning, PCR fusion (Hobert, 2002), and Gibson cloning (Gibson, 2011), we created constructs in which *hlh-2*prox drives expression of a bHLH gene associated with specification or terminal features of a different regulatory cell, in an operon with SL2::2xNLS-mCherry so that successful expression of the untagged bHLH protein could be assessed by mCherry. We used the *gpd-2* SL2 sequence from pXW09 (Wei et al., 2012). All such constructs have the form *hlh-2*prox::*bhlh*::*sl2*::*mcherry::unc-54* UTR, and are abbreviated hereafter as *hlh-2*prox*::bhlh*.Furthermore, all were created in a miniMos vector (pCFJ910) (Frokjaer-Jensen et al., 2014) that would allow for single-copy insertion or analysis in conventional complex arrays.

pHL3: the *bHLH* insert is *lin-32* cDNA obtained from Douglas Portman (Portman & Emmons, 2000).

pHL5: the *bHLH* insert is *hlh-8* cDNA provided by Marian Walhout (Grove et al., 2009)

.

<u>*lag-2-gfp* fosmid-based reporters</u>

To construct the *lag-2-gfp* fosmid reporter, *gfp* was recombineered into the fosmid WRM0620cE03 to stay in frame at the carboxy terminus of *lag-2*; we followed the standard recombineering protocol (Tursun et al., 2009) but with a five-fold increase in PCR product to increase the chances of a recombination event, creating fosMS3.

Generation of transgenes

Generation of *arEx2194*: this transgene marks hDTCs. It is a complex array derived from pMS2, a construct containing a 5.2 kb fragment of the 5' flanking region of *hlh-2* (Sallee et al., 2015). pMS2 (1 ng/µL) was coinjected with pBX [*pha-1*(+)] at 1 ng/µL (Granato et al., 1994), pJL17 (*ttx-3*p::*mCherry*) at 2 ng/µL, and PvuII-digested N2 genomic DNA at 50 ng/µL into *pha-1(e2123ts)* hermaphrodites. Individual injected *pha-1(e2123)* adult hermaphrodites were raised at 15°C for 3 days, and then shifted to 22°C for four days. Independent transgenic strains were isolated (maximum of one per P0). The strain was subsequently kept at 22°C or 25°C to maintain selective pressure to retain the array, which rescues the *pha-1(e2123ts)* lethality.

Generation of transgenes for bHLH ectopic expression: *arTi148*[*hlh-2*prox::*lin-32*] was created by injecting pHL2 into PD4443 (*ccIs4443[arg-1*p::*gfp]*) hermaphrodites, followed by the standard miniMos strain generation protocol (Frokjaer-Jensen et al., 2014). *arEx2500[hlh-2*prox*::hlh-8]* was created by injecting ScaI-digested pHL5 into PD4443 hermaphrodites at 1ng/µL with PvuII-digested N2 genomic DNA at 50 ng/µL and ScaI-digested pCW2 (*ceh-22p*::*gfp*) at 1 ng/µL.

Generation of *arEx2499*[*lag-2-gfp* fosmid]: NotI-digested fosMS3 was injected at 15 ng/µL with Sca-I-digested pBX (*pha-1*(+)) and pCW2 (*ceh-22*p:*gfp*) at 1ng/µL and PvuII-digested N2 genomic DNA at 50 ng/µL into *pha-1(e2123ts*) hermaphrodites. The strain carrying *arEx2499* was established using *pha-1* rescue as described above.

RNAi

RNA interference (RNAi) was performed by feeding bacteria expressing double-stranded RNA specific to a single gene (Kamath et al., 2003) to L1 or L2 worms. We used published RNAi

clones: *hlh-2* (Karp & Greenwald, 2003), 30 sequenced bHLH genes from the Ahringer library (Kamath et al., 2003), and RNAi clones for the remaining eleven bHLH genes (Sallee et al., 2015). RNAi plates were made as previously described: NGM plates with 6mM IPTG and 100uM carbenicillin with 75uL of overnight RNAi bacterial culture (2xYT with 50uM ampicillin). Embryos were collected from gravid hermaphrodites via treatment with a bleach/sodium hydroxide solution, followed by washing in M9, and then embryos hatched on the RNAi plates at 25°C (for L1 RNAi) or grown first on seeded plates and then placed on RNAi plates (for L2 RNAi).

For the hDTC functional screen, L1 worms of strain GS7535[*hlh-2(bx108); arIs222[lag-2*p*::2Xnls-tagrfp]; nre-1(hd20) lin-15b(hd126)* were placed on RNAi bacteria and scored approximately 40 hours later, when *lacZ* negative control larvae are at the early L4 larval stage with vulval invagination is underway and the hDTCs having completed both turns. RNAi targeting *lacZ* was used as a negative control and RNAi targeting *hlh-12* was used as a positive control for every experiment. When present, hDTCs were identified by morphology and by expression of the marker *arIs222[lag-2*p::*tagrfp]*. Although Class II bHLH genes are the most likely to encode dimerization partners for the sole Class I bHLH gene *hlh-2* (Massari & Murre, 2000; Grove et al., 2009), we included all 41 *C. elegans* genes that encode proteins with predicted bHLH domains (other than *hlh-2*) in the screen.

For scoring the requirement of bHLH genes for niche marker gene expression, L2 worms of strain GC956, GS8185 or GS8605 were placed on RNAi bacteria. For GC956 and GS8185, animals were staged based on L2 gonad morphology (gonad elongation underway) and for the Rol phenotype conferred by the transgene, which is not apparent in the L1 stage but is evident in the L2 stage. For GS8605 animals were staged by L2 gonad morphology and if *lag-2* was still expressed in the two α cells, in order to maximize exposure of hDTCs to RNAi treatment. L2

larvae were washed with M9 and transferred onto RNAi plates, and hermaphrodites were scored approximately 40 hours post-embryo collection, between late L3 and mid L4.

When we tested the ability to use RNAi for a functional screen of male regulatory cells, using GS7859[*hlh-2(bx108); arIs222*[*lag-2*p::*tagrfp*]*; ccIs4443*[*arg-1*p::*gfp*] *him-5(e1490); nre-1(hd20) lin-15b(hd126)*], we found that mDTCs are refractory to RNAi (Fig. S3E) and therefore instead relied on the bHLH fosmid expression screen to identify bHLH genes in male regulatory cells.

bHLH fosmid expression screen

Strains carrying translational reporters were available for 17/18 Class II bHLH genes (all but *hlh-32*, as no translational reporter was available). The translational reporters contain a large portion of the genomic context of each gene, with all but HLH-16 being fosmid-based (Sallee et al., 2015). We examined one transgenic strain for HLH-8, HLH-16, and LIN-32 (integrated), and two independent transgenic strains for the remaining 14 bHLH genes. We visually screened these strains for expression in the hDTC, LC and mDTC in all larval stages (Fig. S1). Males were generated by crossing wild-type N2 or *pha-1(e2123)* males to hermaphrodite strains carrying fosmid reporter transgenes, except in the case of *otIs594[lin-32-gfp* fosmid*]*, where the allele *him-5(e1490)* was used to increase the frequency of male self-progeny.

Expression was examined in more detail for the *hlh-2-gfp* fosmid (*arEx2028*), the *lin-32-gfp* fosmid (*otIs594*), the *hlh-12-gfp* fosmid (*arIs235*) the *hlh-8-gfp* fosmid (*wgIs74*), and the *hlh-3-gfp* fosmid (*otIs648*), with a minimum of 10 larvae scored for expression at each stage for each reporter (Fig. S1). All scoring was performed at the 63x PlanApo objective on a Zeiss Axio Imager D1 microscope with a Zeiss AxioCam MRM camera, with expression considered “ON” if

fluorescence was visible at a 500 ms exposure time for GFP.

Assessing phenotypes

hDTC. Leader defect = elongated gonad without completion of both turns by the time vulval invagination was occurring. Leader failure = no evidence of gonad elongation. Niche defect = no expression of *naEx156[apx-1*p*::gfp].* Specification defect: concomitant leader failure and lack of niche marker expression. The niche marker *arIs222*[*lag-2*p::*tagrfp]* was used in the functional RNAi screen and both *arIs222* and *arEx2194[hlh-2*p::*gfp*] were used for other analyses. We used *lag-2*p::*tagrfp* as a marker for hDTC specification rather than for niche function, because *lag-2* expression does not depend on *hlh-2* in a specified hDTC (Fig. S2B).

mDTC. Niche defect = no expression of *ccIs4443[arg-1*p*::gfp].* Specification was assessed using *otIs502[yfp-hlh-2* fosmid*].*

LC. Leader defect = elongated gonad without dorsal and/or posterior turn. Leader failure = swollen gonad due to continued germline proliferation with no evidence of gonad elongation (no outgrowth). Specification was assessed using *arIs222*, but see text for further discussion.

bHLH mutant analysis

In many cases, worms were synchronized to facilitate obtaining large numbers of animals at the appropriate stage. Two methods were used: collecting embryos from gravid hermaphrodites via treatment with a bleach/sodium hydroxide solution followed by washing in M9, or having gravid hermaphrodites lay eggs for two hours, and then determining the appropriate amount of time needed to achieve the desired stage.

*hDTCs***.** Mutant hermaphrodites were grown as homozygotes, except for double mutants with *hlh-12(tk68)* due to sterility. In those cases, *hlh-12(tk68)* homozygotes segregated from

heterozygous mothers, with *hlh-12(tk68)* balanced by either *arIs51* or *oxTi414*. Mutants were grown at 25°C and examined at the L4 stage for evidence of defective hermaphrodite distal tip cell (hDTC) leader function and specification (Fig. 2A,B). Images were taken using the 40X PlanNeo objective of a Zeiss Axio Imager D1 microscope and a Zeiss AxioCam MRm camera. Expression of markers was determined at a 100 ms GFP and 50 ms RFP exposure time was scored as "ON."

To score for niche defects, *naEx156[apx-1*p::*gfp; rol-6*(*su1006*)*]* expression was scored as a proxy for hDTC niche function (Fig. 2C, Fig. S3C). To bypass a requirement for hDTC specification in the L1 stage (Karp & Greenwald, 2004), hermaphrodites were treated with *bhlh* RNAi beginning in the L2 stage. GC956 [*naEx156*] or GS8185 [*lin-32(tm1446); naEx956*] embryos were collected by treating approximately 200 gravid array-carrying hermaphrodites with a bleach/sodium hydroxide solution, and pipetted onto NGM plates seeded with OP50 at 25°C. Affected hDTCs show abnormal gonad elongation on *hlh-2(RNAi)* and *hlh-12(RNAi)*, and were scored for GFP expression using the 63x PlanApo objective of a Zeiss Axio Imager D1 microscope. hDTCs were identified by morphology and position using DIC, and GFP expression scored as "ON" if it was visible with a 500 ms exposure time.

mDTCs. Strains were grown at 25°C and males were imaged at either L2 or L4 (as indicated). As the mutants analyzed for mDTC phenotypes do not have defective gonad arm outgrowth, males were staged based on gonad elongation. GFP expression was scored as "ON" if it was visible by eye, and images were taken at a 10 ms exposure time. All bHLH single and double mutants scored for mDTC phenotypes were maintained and analyzed as homozygotes (Fig. 3B). We observed that *hlh-8(nr2061)* or *hlh-2(bx108)*; *hlh-8(nr2061)* males frequently did not survive past the L2 stage, as determined by the extent of LC-directed gonad elongation; we thus scored all males carrying the *hlh-8(nr2061)* allele at the L2 stage only (Fig. 3B).

LC. Strains were grown at 25°C and scored as described above.

Reprogramming experiments

"Subtraction" experiment. GS8628 and GS8629 were synchronized at 25°C using the egg-laying method described above, and scored approximately 34-36 hours later, during the L3 stage. Due to the completely penetrant migration defect seen in *hlh-3(0)* animals, gonad extension in the wild-type control was used to stage animals, and only *lag-2*p::*tagrfp*-expressing cells were scored for *arTi22* expression at a 500ms exposure for GFP.

To verify that the somatic primordium forms normally in *hlh-3(0)* males, GS7876 and GS8008 animals were synchronized using the egg-laying method at 25°C and scored during late L1, just after LC specification. In *hlh-3(0)* and *hlh-3(+)* male larvae, we observed *arIs222[lag-2*p::*tagrfp]* expression in four anterior somatic gonad cells, consistent with the LC and VDs as in wild-type (Fig. S4A).

"*Addition" experiment.* Embryos were collected using the bleach/sodium hydroxide protocol described above, grown at 25°C, and scored during late L2 for *arg-1*p::*gfp* expression, using the 63x PlanApo objective of a Zeiss Axio Imager D1 microscope. α and β cells with ectopic bHLH expression were identified by position and by the presence of mCherry, which is co-expressed with the bHLH genes from *hlh*-2prox; *arg-1*p::*gfp* expression was scored as "ON" if visible at a 25ms exposure time. mCherry was always observed in only four cells.

As shown in Figure 4B, +LIN-32 +HLH-8 hermaphrodites also display mDTC-like processes towards the germline in α and β cells expressing *arg-1*p::*gfp*. To visualize these processes, we used a spinning disk confocal microscope (Carl Zeiss) and captured images using the 100x objective. These hermaphrodites were also scored for functional AC defects (Fig. S4C-G).

In a wild-type hermaphrodite, the AC induces vulval development and invagination in the L3 stage, and joins the uterine-seam cell (utse) in the L4 stage. Hermaphrodites were raised at 25°C and L4 larvae were examined for defects in these AC-dependent processes by DIC at 63x on the Zeiss Axio Imager D1 microscope. As *hlh-2*prox does not drive expression in the hDTCs, gonad elongation was predicted to be normal and used to stage animals. Hermaphrodites were scored as having abnormal vulval induction if the vulval precursor cell P6.p had divided only once by the time both gonad arms had reflexed, as identified using DIC, and as having abnormal vulval invagination if there was no evidence of invagination by the time both gonad arms had reflexed or if defects such as multiple connected invagination sites were seen (Fig. S4C, S4D). Animals were also scored for abnormal vulval morphology, defined as large "blips" at the vulva site (Fig. S4G). Fusion of the AC to the utse was considered abnormal if the AC nucleus was visible above the invaginated vulva at the time of normal utse formation (Fig. S4E, S4F).

To verify that the somatic primordium forms normally in the +LIN-32 +HLH-8 hermaphrodites, GS8726 and GS8630 animals were synchronized using the bleach/sodium hydroxide protocol described above, grown at 25°C, and scored during late L1/early L2 for cells expressing either GFP-HIS-58 or mCherry showing the stereotypical position of the α and β cells ([19, 51], Fig. S4B).

In contrast to most Class II bHLH proteins, which form obligate dimers with HLH-2/Da/E proteins [16, 17], HLH-8/Twist has been proposed to function as either homodimers or heterodimers [43]. Since HLH-2 is stable only in the AC [18], but *hlh-2*prox drives ectopic expression of HLH-8 in all four cells with AC potential, the observation that *arg-1* is expressed only in the AC when HLH-8 alone is ectopically expressed implies that HLH-8:HLH-2 heterodimers (rather than HLH-8 homodimers) drive *arg-1* expression. Furthermore, the

observation that all four cells express *arg-1* in +LIN-32+HLH-8 hermaphrodites is further support for the interpretation that they are transformed into mDTCs, where HLH-2 should be stable.

Quantification and statistical analysis

When comparing two genotypes for the frequency of two outcomes, a two-tailed 2x2 Fisher's exact test was used. This test was used in Fig. 2C. When comparing two genotypes for the frequency of three different outcomes, a two-tailed 2x3 Fisher's exact probability test was used. This test was used in Figs. 2A, 2B, and 3. For all cases, differences were considered significant if the p-value is less than 0.01.

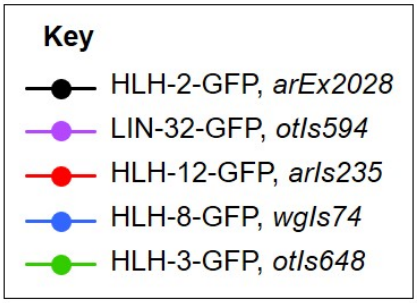
Table 2.1. Key resources table

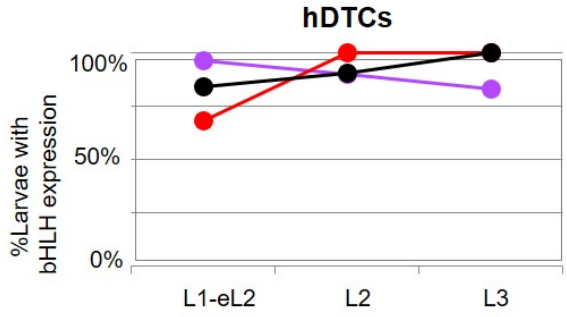
REAGENT or RESOURCE	SOURCE	IDENTIFIER
C. elegans Strains		
GC956: *naEx156*[*apx-1*p::*gfp*]	[23]	WB Strain: GC956
GS4581: *arIs51[cdh-3*p::*gfp] IV*	[5]	
GS5004: *hlh-2(bx108ts) I*	This paper	
GS7055: *hlh-2(bx108) I; lin-32(tm1446) X*	This paper	
GS7428: *pha-1(e2123) III; arEx2028[hlh-2-gfp* fosmid*]*	[18]	
GS7535: *hlh-2(bx108) I; arIs222[lag-2*p::*2xnls-tagrfp] V; nre-1(hd20) lin-15b(hd126) X*	[18]	
GS7687: *hlh-12(tk68/arIs51[cdh-3*p::*gfp]) IV; lin-32(tm1446) X*	This paper	
GS7690: *hlh-2(bx108) I; hlh-12(tk68)/arIs51[cdh-3*p::*gfp] IV*	This paper	
GS7726: *arIs235[hlh-12-gfp* fosmid*]; him-5(e1490) V*	This paper	
GS7732: *otIs594[lin-32-gfp* fosmid*]; him-5(e1490) V*	This paper	
GS7826: *pha-1(e2123) III; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V; arEx2194[hlh-2*(5.2kb)p::gfp*]*	This paper	N/A
GS7827: *pha-1(e2123) III; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490 V); lin-32(tm1446) X; arEx2194[hlh-2*(5.2kb)p::gfp*]*	This paper	N/A
GS7851: *qIs90[ceh-22*p::*venus]; hlh-12(tk68) IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V*	This paper	
GS7853: *pha-1(e2123) III; hlh-12(tk68) IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490 V); arEx2194[hlh-2*(5.2kb)p::gfp*]*	This paper	N/A
GS7854: *pha-1(e2123) III; hlh-12(tk68)/oxTi414[eft-3*p::*mcherry] IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490 V); lin-32(tm1446) X; arEx2194[hlh-2*(5.2kb)p::gfp*]*	This paper	N/A
GS7859: *hlh-2(bx108) I; ccIs4443[arg-1*p::*gfp] IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V; nre-1(hd20) lin-15b(hd126) X*	This paper	
GS7874: *lin-32(tm1446)X*	This paper	
GS7876: *ccIs4443[arg-1*p::*gfp] IV*; *arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V*	This paper	
GS7877: *ccIs4443[arg-1*p::*gfp] IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V; lin-32(tm1446) X*	This paper	
GS7878: *hlh-2(bx108) I; ccIs4443[arg-1*p::*gfp] IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V*	This paper	

GS7879: *hlh-2(bx108) I; ccIs4443[arg-1*p::*gfp] IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V; lin-32(tm1446) X*	This paper	
GS7883: *otIs648[hlh-3-gfp* fosmid*]; histone-rfp*	This paper	
GS8008: *hlh-3(tm1688) II; ccIs4443[arg-1*p::*gfp] IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V*	This paper	
GS8185: *lin-32(tm1446) X; naEx156*	This paper	
GS8196: *arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V; naEx156[apx-1*p::*gfp]*	This paper	
GS8518: *arTi148[hlh-2prox::lin-32cDNA::sl2::mcherry]; ccIs4443[arg-1*p::*gfp] IV*	This paper	
GS8605: *pha-1(e2123) III; arEx2499[lag-2-gfp* fosmid*]*	This paper	
GS8606: *ccIs4443[arg-1*p::*gfp] IV; arEx2500[hlh-2prox::hlh-8cDNA::sl2::mcherry]*	This paper	
GS8614: *hlh-2(bx108) I; ccIs4443[arg-1*p::*gfp] IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V; hlh-8(nr2061) X*	This paper	
GS8615: *ccIs4443[arg-1*p::*gfp] IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V; hlh-8(nr2061) X*	This paper	
GS8626: *him-8(e1489) IV; otIs502[yfp-hlh-2* fosmid*]*	This paper	
GS8627: *him-8(e1489) IV; otIs502[yfp-hlh-2* fosmid*]; lin-32(tm1446) X*	This paper	
GS8628: *ccIs4443[arg-1*p::*gfp] IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V; arTi22[hlh-2prox::gfp-his-58]*	This paper	
GS8629: *hlh-3(tm1688) II; ccIs4443[arg-1*p::*gfp] IV; arIs222[lag-2*p::*2xnls-tagrfp] him-5(e1490) V; arTi22[hlh-2prox::gfp-his-58]*	This paper	
GS8630: *arTi148[hlh-2prox::lin-32cDNA::sl2::mcherry]; ccIs4443[arg-1*p::*gfp] IV; arEx2500[hlh-2prox::hlh-8cDNA::sl2::mcherry]*	This paper	
GS8631: *arIs131[lag-2*p::*2xnls-yfp] III; him-8(e1489) IV*	This paper	
N2: +	Caenorhabditis Genetics Center	WB Strain: N2
NF1253: *hlh-12(tk68)IV*	[7]	
OP74: *unc-119(ed3) III; wgIs74[hlh-8-gfp* fosmid*]*	Caenorhabditis Genetics Center	WB Strain: OP74
PD4443: *ccIs4443[arg-1*p::*gfp] IV*	Caenorhabditis Genetics Center	WB Strain: PD4443

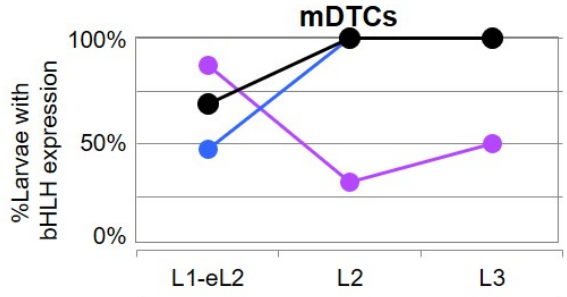
Supplemental information

bHLH Reporter	AC	hDTC	LC	mDTC
CND-1-GFP				
HLH-2-GFP	+	+	+	+
HLH-3-GFP			+	
HLH-4-YFP				
HLH-6-YFP				
HLH-8-GFP				+
HLH-10-YFP				
HLH-12-GFP		+	+	
HLH-13-GFP				
HLH-14-GFP				
HLH-15-GFP				
HLH-16-GFP				
HLH-17-GFP				
HLH-19-YFP				
HLH-31-GFP				
HND-1-GFP				
LIN-32-GFP		+	+	+
NGN-1-GFP				

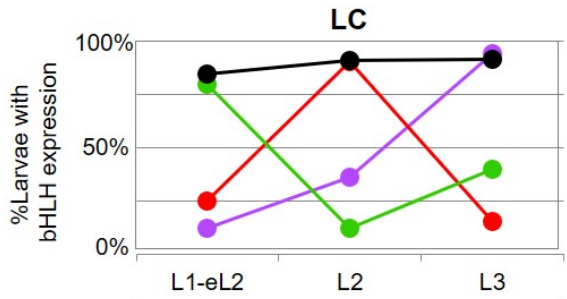




LIN-32-GFP	28/29	18/20	15/18
HLH-12-GFP	17/25	20/20	22/22
HLH-2-GFP	16/19	19/21	20/20

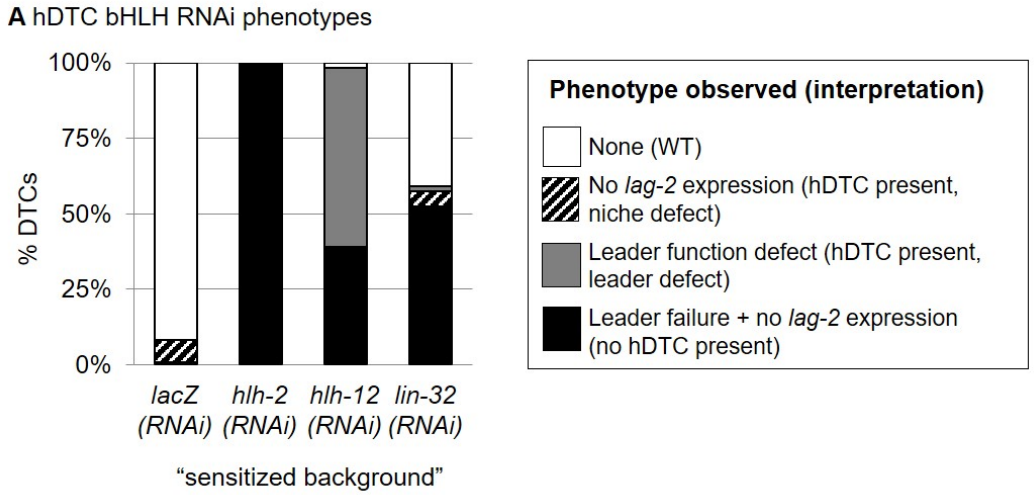


HLH-8-GFP	11/23	15/15	17/17
LIN-32-GFP	14/16	7/22	9/18
HLH-2-GFP	9/13	11/11	12/12



HLH-3-GFP	8/10	2/16	4/10
LIN-32-GFP	2/16	8/22	17/18
HLH-12-GFP	3/13	19/21	2/13
HLH-2-GFP	11/13	10/11	11/12

Figure S2.1. bHLH expression screen and time course in regulatory cells. Related to Figure 1. We previously found that transgenes expressing GFP translational fusions with Class II bHLH proteins are not expressed during anchor cell (AC) development [S1]. For this study, we visually screened the same transgenes for expression in the hDTC, LC, and mDTC in all larval stages. Expression is denoted by green fill and "+"; blank indicates no expression. Expression of all Class II bHLH genes was examined except *hlh-32*, for which no translational reporter was available. See STAR Methods ("bHLH fosmid expression screen") for scoring details. Reporters that were expressed in the LC, the hDTCs, or the mDTCs in the initial expression screen were scored in more detail over the course of larval development. The tables below show the number of GFP(+) worms out of the total number scored at each larval stage and for each regulatory cell. The data for each table is displayed above it as a graph.



B *hlh-2(L2-RNAi)*: *hlh-2* is not required for *lag-2* expression in hDTCs after hDTC specification

	hlh-2 (L1-RNAi)	*hlh-2* (L2-RNAi)
arIs222 (*lag-2p::tagrfp*)	0% (42)	100% (48)
arEx2499 (*lag-2-gfp* fosmid)	0% (40)	100% (46)

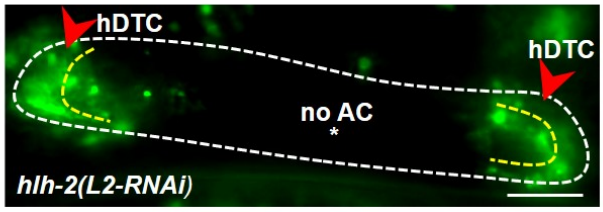


Figure S2.2. bHLH genes are required in the hDTC for specification and terminal features, but not *lag-2* expression. Related to Figure 2.

A. We quantified the phenotypes for the candidates identified in the hDTC RNAi screen (see Table S1) in the same "sensitized" background GS7535 [*nre-1(hd20) lin-15b (hd126)]*, to sensitize for RNAi [S2], and *hlh-2(bx108ts)*, which compromises the ability of HLH-2 to form dimers (Portman & Emmons, 2000). *lag-2::tagRFP (arIs222)* marked the hDTCs. Treated individuals were scored during L4 for defects in gonad arm migration and/or *lag-2* expression.

B. If *hlh-2(RNAi)* is performed in the L1 stage, the prospective hDTCs are not specified as such (Karp & Greenwald, 2004; Chesney et. al, 2009), and therefore *lag-2* reporters and other markers for hDTCs are not expressed. If RNAi is performed in the L2 stage, the hDTCs are specified normally and display the expected defect in outgrowth of the gonad arms, indicating that RNAi is effective; however, two different *lag-2* reporters are expressed, including a fosmid reporter that should contain all essential regulatory sequences, suggesting that *hlh-2* is not required for *lag-2* expression in specified hDTCs. Because *lag-2* does not depend on *hlh-2* for expression in a specified hDTC, we were able to use the presence of *lag-2* as a marker of a specified hDTC for our genetic analyses. The photomicrograph shows representative LAG-2-GFP (*arEx2499*) fosmid expression in a late L3 hermaphrodite treated with *hlh-2*(*L2-RNAi).* AC specification failed (asterisk) and both hDTCs exhibit leader failure but still clearly express LAG-2-GFP (arrowheads). White dotted line outlines gonad, yellow dotted lines outline hDTCs, and scale bar is 10 μm. See STAR Methods ("bHLH RNAi") for experimental details.

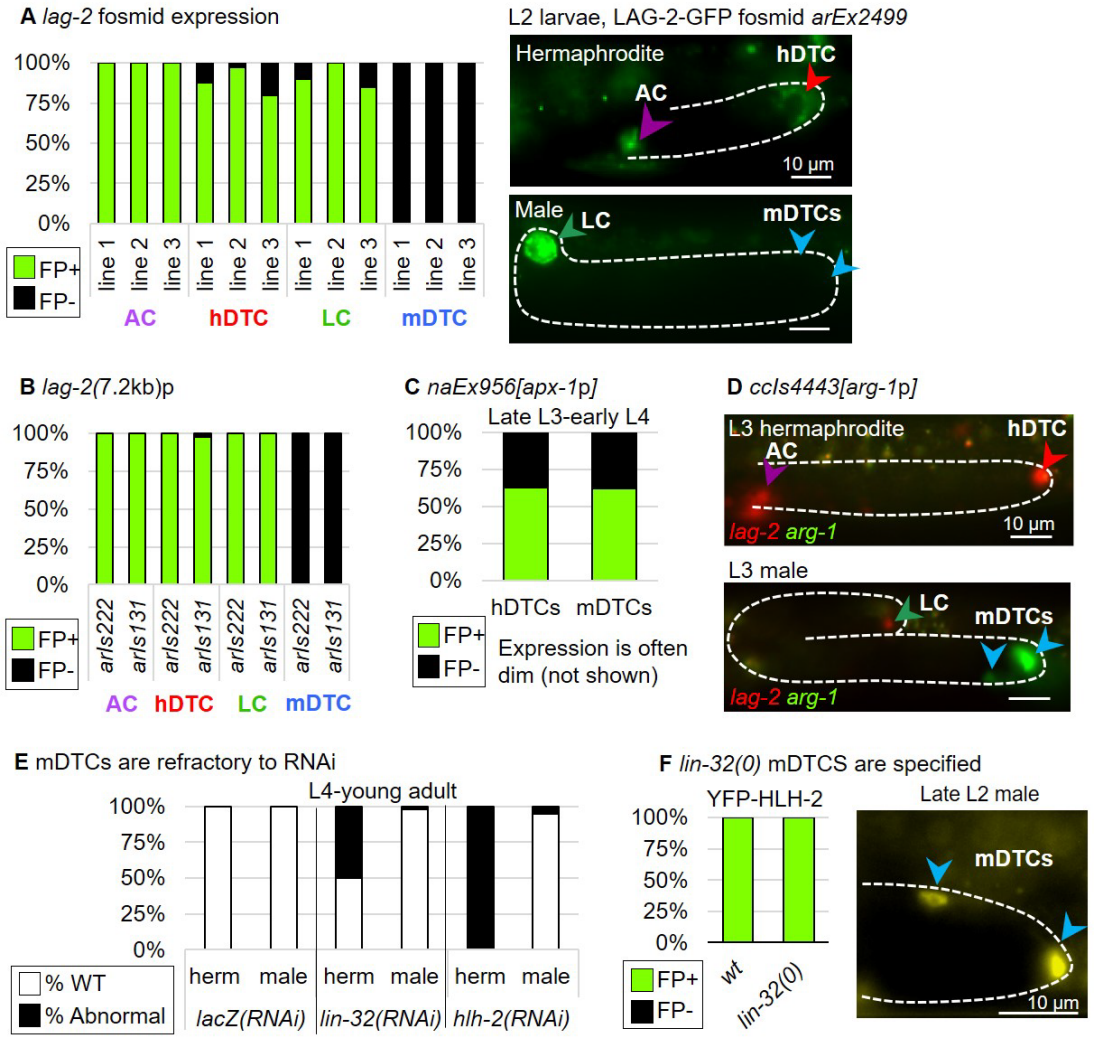


Figure S2.3. Sexual dimorphism of niche ligands for GLP-1/Notch, insensitivity of mDTCs to RNAi, and mDTC specification in a *lin-32(0)* background. Related to Figure 3.

A. *lag-2* has been considered a niche ligand in the mDTCs based on low-level expression of *qIs57,* which contains a 3.3 kb "promoter" fused to *gfp* (Chesney et al., 2009). However, *qIs57* is not expressed in the AC, so the construct from which it was derived is missing regulatory sequences. Here, we show that three fosmid-based *lag-2* reporters display robust expression in the AC, LC, and hDTCs, but not in the mDTCs. Thus, *lag-2* may not be a bona fide niche ligand in mDTCs.

B. Two independent *lag-2* transcriptional reporters with 7.2kb of 5' *lag-2* sequences driving expression of YFP (*arIs131*) and 2xNLS-Tag-RFP (*arIs222*), have the same expression as fosmid reporters.

C, D. By contrast, reporters for the ligand genes *apx-1* and *arg-1* are expressed in mDTCs, suggesting they are true male niche ligands.

E. mDTCs are refractory to RNAi, even in the sensitized genetic background GS7859 (see Key Resources Table, Fig. S2A legend). Males did not show a defect in *arg-1* expression in mDTCs, even when RNAi in sibling hermaphrodites caused hDTC specification to fail.

F. *lin-32(0)* mDTCs display strong YFP-HLH-2 (*otIs502*) expression and have normal shape, size, and position within the gonad. Thus, *lin-32* is required for *arg-1* expression (Fig. 3B), but not mDTC specification. Dotted white line outlines the gonad, blue arrowheads mark the mDTCs.

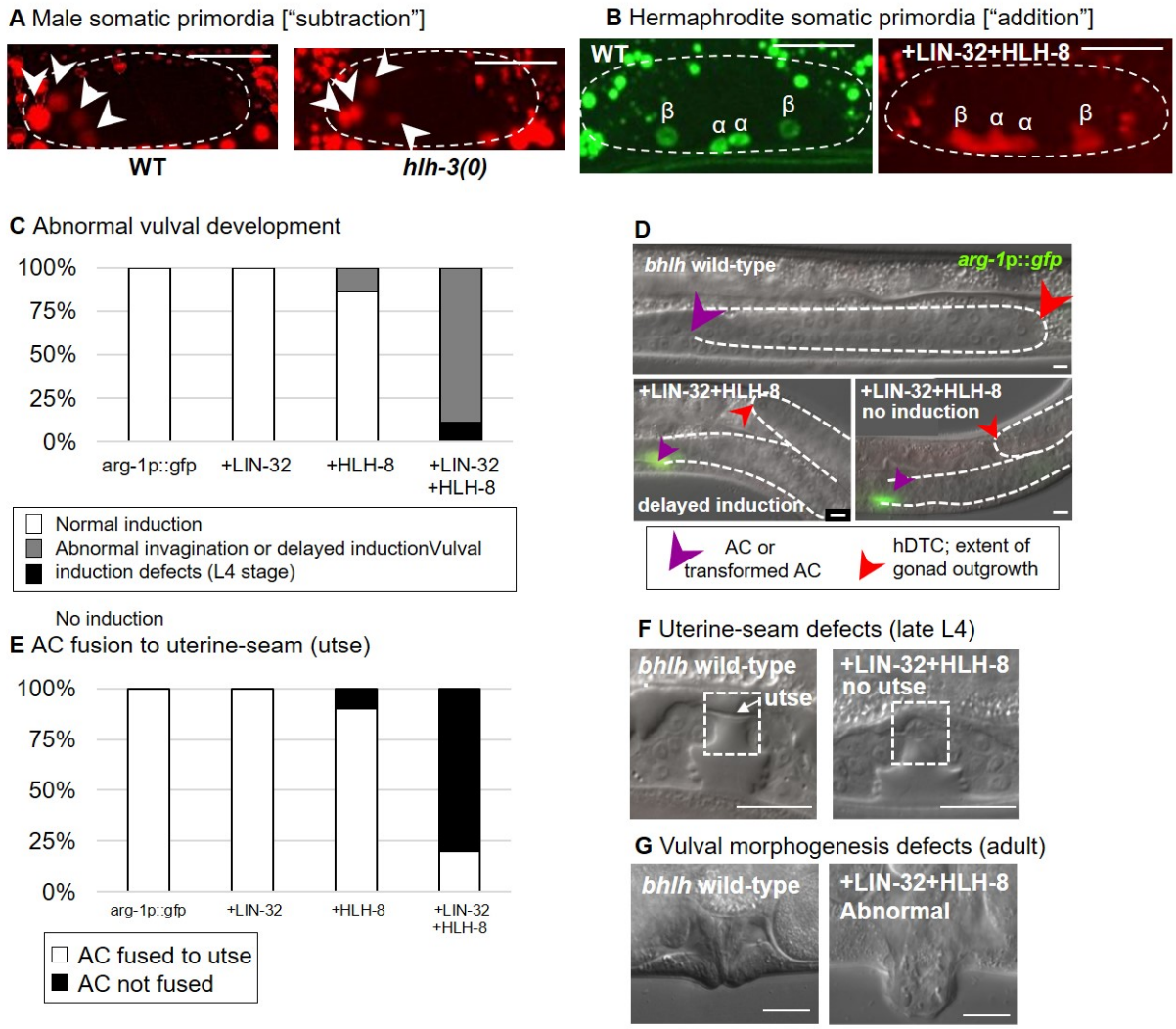


Figure S2.4. Gonad primordium formation in reprogramming experiments and compromised terminal AC function in LIN-32+HLH-8 "addition" hermaphrodites. Related to Figure 4.

A, B. Gonad primordia (outlined in dotted white lines) form normally even when regulatory cells are reprogrammed b y altering their bHLH gene complement, suggesting the gonad lineages are normal. Scale bars are 10 µm. A: Both WT and *hlh-3(0)* males (n=20 for each) express *lag-2*p::*tagrfp* in the prospective LC and VDs in the L1 (white arrowheads). B: The a and b cells are marked in WT with *hlh-2*prox*::gfp-his-58* (n=22); in +LIN-32+HLH-8, they are marked by mCherry expressed from the *hlh-2*prox::*lin-32*::sl2::*mcherry* and *hlh-2*prox::*hlh-8*::sl2::*mcherry* transgenes (n=21).

C, D. Vulval defects. Induction of vulval development and organization of vulval morphogenesis are critical functions of the AC. In wild type, 22 vulval cells are present in a stereotypical pattern

and position in the L3 stage, prior to hDTC-guided gonad arm reflexion. Vulval development is abnormal (delayed or disorganized) when *hlh-2*prox is used to express both LIN-32 and HLH-8 ectopically in the AC, as readily apparent when L4 hermaphrodites are scored, and the gonad arms have reflexed. C: "+ (bHLH)" genotypes include *ccIs4443[arg-1* p::*gfp]*. D: Dotted white lines outline the gonads; scale bars are 10 μm.

E, F. Uterine defects. The AC organizes uterine cell fate and morphogenesis. The AC fuses to the uterine seam cell (utse) in the late L4 stage (arrow). Utse fusion frequently fails when LIN-32 and HLH-8 are ectopically expressed in the AC. Graph shows percent of L4 hermaphrodites with normal or failed AC-utse fusion, and F shows representative photomicrographs.

G. This common abnormal vulval morphogenesis defect may reflect failure of AC-utse fusion or other aspects of defective AC function. F, G: scale bars are 10 μm.

Table S2.1. Functional screen of bHLH genes for hDTC defects. Related to Figures 1 and 2. *Genetic background*: *hlh-2*(*bx108*) sensitizes worms to redundancy of bHLH partners for HLH-2 & Emmons, 2000), *arIs222* (*lag-2*p::*tagrfp;* Sallee & Greenwald, 2015) marks the hDTC, and *nre-1(hd20) lin-15b(hd126)* sensitizes somatic tissue to RNAi (Schmitz et al., 2007). Defects were scored as described in the STAR Methods. n = 20 for all experiments. Bold genes are significantly different than controls, **p<0.001 by two-tailed 2x3 Fisher's exact probability test. *Positive control:* a leader defect caused by *hlh-12(RNAi)* [S6]. *Negative control: lacZ(RNAi*).

	RNAi	0 DTC	1 DTC	2 DTC
Neg. control	*lacZ*	0%	5%	95%
Class II	*cnd-1*	0%	0%	100%
	hlh-3	0%	0%	100%
	hlh-4	0%	5%	95%
	hlh-6	0%	0%	100%
	hlh-8	0%	0%	100%
	hlh-10	0%	0%	100%
	*hlh-12***	95%	5%	0%
	hlh-13	0%	0%	100%
	hlh-14	0%	0%	100%
	hlh-15	0%	5%	95%
	hlh-16	0%	0%	100%
	hlh-17	0%	0%	100%
	hlh-19	0%	5%	95%
	hlh-31/32	0%	0%	100%
	hnd-1	0%	0%	100%
	*lin-32***	45%	40%	15%
	ngn-1	0%	5%	95%
Class III	*hlh-30*	0%	5%	95%
	sbp-1	0%	0%	100%
Class IV	*mdl-1*	0%	5%	95%
	mml-1	0%	10%	90%
	mxl-1	0%	0%	100%
	mxl-2	0%	0%	100%
	mxl-3	0%	0%	100%
Class VI	*hlh-25*	0%	5%	95%
	hlh-26	0%	0%	100%
	hlh-27	0%	5%	95%
	hlh-28	0%	0%	100%
	hlh-29	0%	5%	95%
	lin-22	0%	0%	100%
	ref-1	0%	0%	100%
Class VII	*aha-1*	0%	0%	100%
	ahr-1	0%	0%	100%
	cky-1	0%	0%	100%
	hif-1	0%	5%	95%
	hlh-33	0%	5%	95%
	hlh-34	0%	0%	100%

Class unknown	*hlh-1*	0%	5%	95%
	hlh-11	0%	5%	95%
	unc-3	0%	5%	95%

Chapter 3: Further investigations into the bHLH code

Introduction

In Chapter 2, I used *hlh-2prox*, a regulatory element that drives specific expression in the four cells that initially have the potential to be an AC ("pro-AC"), their parents, and the differentiated AC (Sallee and Greenwald, 2015) to express LIN-32 and HLH-8, the mDTC Class II code genes, and observed transient reprogramming of the four pro-ACs into mDTCs in the L2 stage. The pro-AC→mDTC transformation occurs despite differences in sex, position, and lineage inputs; therefore, I expected to be able to effect pro-AC→hDTC and pro-AC→LC transformations in the same manner, as the first transforms between cells that differ in only position and lineage and the second transforms between cells that differ in sex only.

As described in Chapters 1 and 2, cell fate reprogramming can occur in many ways in both natural and artificial settings (reviewed in Rothman and Jarriault, 2019). However, one of the most important hallmarks for determining if a reprogramming event has taken place is the permanence of the adoption of the new cell fate. Notably, there are examples of both natural and artificial full and permanent cell fate reprogramming in *C. elegans* (see Chapter 1.4 for further discussion). In Chapter 2, I showed that the effected pro-AC→mDTC transformation was strongest in the L2 stage; while marker expression continued through L4, morphological changes did not, suggesting that the cell was losing its adopted mDTC fate over time. Similarly, I found below that adding the hDTC or LC bHLH complements to the pro-ACs resulted in a transient change in cell fate, and so describe these cases as pro-AC transformations instead of pro-AC reprogramming *per se*. I discuss some potential barriers to full, permanent cell fate reprogramming in this context here and in Chapter 5.

While I was able to generally effect pro-AC→hDTC or pro-AC→LC transformation in accordance with the bHLH code hypothesis, there were some unexpected features with the pro-AC→LC context. An analysis of endogenously-tagged bHLH genes performed by Sarah Finkelstein led to a new understanding of the expression patterns of these proteins during the specification of each regulatory cell fate, indicating that their expression and function during these different cell fate specification events may be more dynamic than our analysis of the fosmid reporters in Chapter 2 suggested. In addition, I provide separate genetic evidence for the function of LIN-32 in LC fate specification. Taken together, these results suggest that, while it is possible to transform the four pro-ACs into any of the non-AC regulatory cell types by expression of the appropriate bHLH complement, each context demonstrates unique features which could come from different requirements for bHLH expression in the endogenous setting.

Materials and methods

C. elegans genetics (see Supplemental Figure 1C for simplified marker expression)

All strains were grown under standard conditions at 15, 20, or 25°C as previously described (Brenner, 1974). Unless stated otherwise, all strains were scored at 25°C. In addition to *5(e1490),* used to generate self-progeny males, the following alleles and transgenes were used for this study. Note that expression is described in terms of the somatic gonad only:

arIs51[cdh-3p::gfp] (Karp & Greenwald, 2003) is expressed in the AC after specification in late L2 through L3, and later in the developing uterine seam cell in L4.

arIs222[lag-2p::2xlns-tagRFP] (M. D. Sallee & Greenwald, 2015) is expressed in all somatic gonad regulatory cells except for the mDTCs, as well as P6.p and its descendants. I will refer to this marker below as *lag-2p::rfp*.

qIs90[ceh-22bp::venus] (Lam et al., 2006) expresses in the hDTCs and their sisters, and the mDTCs.

hlh-12(1080)p (Tamai & Nishiwaki, 2007a), used in *arEx2546, arEx2547,* and *nsIs497* (this study; Shai Shaham, personal communication), is expressed in the hDTCs and LC beginning in L2. *arEx2546* and *arEx2547* are *hlh-12(1080)p::gfp-his-11,* and were both made in this study from pCT23 (Claudia Tenen, personal communication). *nsIs497* is *hlh-12(1080)p::venus*. *nsIs497* also contains *nhr-67(2kb)p::mcherry,* which expresses in the AC beginning in L3 and the LC competence group and later LC beginning in late L1.

arTi112[ckb-3p::mcherry-his-11] (Attner et al., 2019) expresses in the somatic gonad beginning in Z1 and Z4, and remains on through at least the primordium stage in all somatic gonad cells.

bHLH alleles (see Chapter 2): *hlh-3(tm1688)* and *lin-32(tm1446)* both contain large deletions including the bHLH domain, making them likely null alleles. *hlh-2(bx108)* is a hypomorphic temperature-sensitive allele that has weakened dimerization ability at the restrictive temperature (Portman & Emmons, 2000b).

bHLH ectopic expression alleles (Sallee, Littleford, & Greenwald, 2017 and this study): *arTi148* (made from pHL3) expresses LIN-32, *arTi275* and *arTi276* (pHL19) express HLH-3, and *arSi45* (pHL53) expresses HLH-12. See below for transgene construction details.

bHLH endogenously-tagged alleles: *hlh-2(623)* is *gfp::hlh-2* from Attner et al., 2019. *hlh-3(vlc28)* is *hlh-3::mNeonGreen* from Lloret-Fernández et al., 2018. All other bHLH endogenously-tagged alleles are new to this paper. See Supplemental Figure 3.2A for an example of their structure. The following new alleles were used:

hlh-8(ar644): gfp::hlh-8
hlh-12(ar643): gfp::hlh-12

lin-32(ar633): gfp::lin-32
lin-32(ar642): lin-32::gfp

Cloning methods

Where listed, "*hlh-2(prox)*" refers to the *hlh-2(prox)* element described in Sallee and Greenwald 2015, used in combination with a regulatory element "S" to increase basal expression (Fire et al., 1990). "*sl2*" is the splicing acceptor sequence from CEOPX036*,* used here to create a bi-cistronic construct (Wei et al., 2012; see also Chapter 2). For fusion PCR, see Hobert, 2002; for Gibson assembly, see Gibson et al., 2009.

*hlh-2(prox)::hlh-3*cDNA*::sl2::mcherry::unc-54 3'UTR* (pHL19): *hlh-3* cDNA was made by amplifying both *hlh-3* exons and fusing them together using fusion PCR. *sl2::mcherry::UTR* was amplified from pMA108 (Michelle Attner, personal communication). Both pieces were fused together using fusion PCR and inserted using Gibson assembly into pHL1 (se Chapter 2) digested with XmaI/NotI, to create pHL16. The full *hlh-3cDNA::sl2::mcherry::unc-54 3'UTR* piece was amplified from pHL16 and inserted into pHL4 (see Chapter 2) digested with SmaI and NotI, to create pHL19.

*hlh-2prox::hlh-12*cDNA::*sl2::2xtagBFP2::unc-54 3'UTR* (pHL53): *hlh-12* cDNA was ordered from Genewiz (genewiz.com). *hlh-2prox* was amplified from pHL4 (see Chapter 2), fused together with *hlh-12*cDNA using fusion PCR, and inserted using Gibson assembly into SmaI/XbaI-digested pBS SK+ (Mayer, 1995)to create pHL48. The *unc-54 3'UTR* was amplified from pHL4. *2xtagBFP2*, hereafter "*BFP*," was amplified from pZW109 (Dickinson, Pani, Heppert, Higgins, & Goldstein, 2015), with the addition of the first 15bp of the 2xGGGGS linker sequence plus an artificial ochre stop codon to allow for better amplification of the repetitive sequence. The two pieces were inserted into SmaI/XbaI-digested pBS SK+ (Mayer, 1995) using Gibson Assembly to

create pHL50. *sl2* was amplified from pHL3 (Chapter 2), and fused together with *BFP::unc-54 3'UTR* amplified from pHL50 to create *sl2::BFP::unc-54 3'UTR*. *hlh-2prox::hlh-12cDNA* was amplified from pHL48. *hlh-2prox::hlh-12cDNA* and *sl2::BFP::unc-54 3'UTR* were inserted into SpeI/AvrII-digested pZW111 (Dickinson et al., 2015) using Gibson assembly to create pHL53.

hlh-2prox::gfp::bHLH constructs for bHLH protein stability: pMS157 is a vector containing *hlh-2prox::gfp::hlh-8cDNA::unc-54* 3'UTR (M. D. Sallee & Greenwald, 2015). PCR was used to amplify pMS157 without the *hlh-8* cDNA, and to amplify either *hlh-3* cDNA (from pHL19) or *hlh-12 cDNA* (from pHL53). The amplified pMS157 fragment was joined with either *hlh-3* or *hlh-12* cDNA using Gibson assembly to form pHL23 (*hlh-3)* or pHL53 (*hlh-12*).

bHLH CRISPR constructs: Homology repair template constructs were cloned by Sarah Finkelstein using the method described in Dickinson et al. 2015, using pDD282 digested with ClaI/SpeI for N-terminal tags and with AvrII and SpeI for C-terminal tags. The following plasmids were used for repair templates:

hlh-8: pHL46
hlh-12: pHL45
lin-32 N-terminal tag: pHL44
lin-32 C-terminal tag: pHL63

For cloning sgRNAs, I used pJW1285 (Ward, 2014) digested with NdeI and SpeI. I then created two PCR fragments designed to replace all of the missing portion of pJW1285, which overlapped using our sgRNA of choice and omitted the original sgRNA contained in the plasmid. These PCR fragments also included overhangs into the pJW1285 backbone, and I used Gibson assembly to fuse the two fragments back into the digested vector. This resulted in a new vector identical in sequence to pJW1285, with the exception of the new added sgRNA at the site of the original. The following sgRNA vectors were used:

hlh-8: pHL37 (5'), pHL39 (3')

hlh-12: pHL33 (5'), pHL34 (3')
lin-32 N-terminal tag: pHL28 (5'), pHL30 (3')
lin-32 C-terminal tag: pHL65 (5'), pHL67 (3')

Generation of transgenes

All plasmids for injection were either ethanol precipitated or cleaned using the PureLink™ Quick Plasmid Miniprep Kit (ThermoFisher).

Endogenously-tagged bHLH alleles: Endogenously-tagged bHLH alleles were generated in collaboration with Sarah Finkelstein using the SEC protocol (Dickinson et al., 2015), with a few modifications. In general, worms were injected with two separate sgRNAs, each at 25ng/µL, and the repair template at 50ng/µL. Injection mixes also contained pGH8 and pCF90, each at 10ng/µL. P0s were treated with 400 µL 5mg/mL hygromycin (Invivogen) three days after injection. Approximately 100 L1 from plates with homozygous dark rollers were heat-shocked at 34°C for four hours, and lines were established from single non-rolling L4 progeny of heat-shocked parents. See the previous section for plasmids used.

miniMos lines: worms were injected following the standard protocol (Frøkjær-Jensen et al., 2014a)with the following modifications: pCFJ104 was not used, and pCFJ601 was injected at 65ng/µL. Injected P0 were placed at 25°C, and treated with either 500 µL 25mg/mL G418 (GoldBio) or 400 µL 5mg/mL hygromycin three days after injection. Lines were established from single animals surviving heat-shock at 34°C for two hours.

bHLH protein stability arrays: arrays were injected into *pha-1(e2123)*, with the *hlh-2prox::bHLHcDNA* at 1ng/µl, ScaI-digested pBX at 1ng/µL, ScaI-digested pCW2.1 at 1ng/µL, and PvuII-digested N2 gDNA at 50ng/µL (following methods from Sallee & Greenwald, 2015).

Injected P0 were left at 15°C for 3-4 days and then transferred to 25°C. Lines were isolated from single array-carrying F1s.

hlh-12(1080)p::gfp arrays: to rescue wild-type vulval development and generate a larger brood size, strains were treated with RNAi targeting *hlh-12* as described below before scoring. Worms were plated as embryos and left on RNAi for approximately two days, and checked quickly with either scope ("General scoring information," below) to confirm WT vulva formation before injection. Worms were injected with 2ng/µL ScaI-digested pCT23, 20ng/µL pGH8 (Frøkjær-Jensen et al., 2014), and 50ng/µL PvuII-digested N2 genomic DNA. Lines were established from single array-carrying F1.

General scoring information

Unless otherwise noted, I synchronized worms for scoring and raised them at 25°C. To synchronize worm age, we used either egg lays or egg preps. For egg lays, approximately 40 adult hermaphrodites were placed on a seeded plate and allowed to lay eggs for 1-2 hours at 25°C before being removed. For egg preps, worm embryos were collected by treating gravid hermaphrodites with a solution of bleach and sodium hydroxide until bursting. Embryos were then washed with water or M9 and either transferred directly to seeded plates, or left in M9 on a nutator for 24 hours at room temperature to reach L1 arrest before plating.

We used three different microscopes in this chapter. The bulk was scored with either Zeiss Axio Imager D1 microscope with either a 40x or 63x PlanNeo objective and a Zeiss AxioCam MRm camera ("right scope"), or a Zeiss Axio Imager Z1 microscope with either a 40x or 63x PlanNeo objective and a Hamamatsu Orca-ER camera. In addition, we used a spinning disk confocal microscope (Carl Zeiss) with a 40x, 63x, or 200x PlanApo objective.

Scoring of endogenously-tagged bHLH alleles

Worms were synchronized using an egg lay, and scored 14-26 hours (late L1-L2) or 38-32 hours (L3) after adult hermaphrodite removal. All worms were scored using the confocal, with the 488m laser at 46.7% power and at 63x. Either *ckb-3p::mcherry-his-11* or *lag-2p::rfp* were used to identify somatic gonad regulatory cells, and *lag-2p::rfp* was used to stage worms precisely. The following exposure times were used:

hlh-2: 550ms exposure
hlh-3: 550ms exposure
hlh-8: 550ms exposure
hlh-12: 500ms exposure
lin-32 N-terminal tag: 1050ms exposure
lin-32 C-terminal tag: 1050ms exposure

RNA interference treatment

RNAi was conducted by raising worms on feeding bacteria containing the appropriate clone at 25°C until the desired stage. All RNAi clones used came from the Ahringer library (Kamath et al. 2003), transformed into HT115 bacteria for feeding, and RNAi targeting *lacZ* was used as a negative control. RNAi cultures were grown in 2xYT or LB media with 50µg/mL carbenicillin at 37°C overnight, and 70µL of each culture was added to room-temperature NGM plates with added 6uM IPTG and 100uM carbenicillin (hereafter referred to as RNAi plates). Plates were allowed to dry for a minimum of two hours or up to overnight at room temperature before worms were added.

For LC *cdh-3p::gfp* scoring: worms were plated on to RNAi after L1 arrest and scored 18-20 hours after plating for late L2, or 26-30 hours after plating for late L3, using the right scope with the 63x objective. As most genetic backgrounds and RNAi conditions in this experiment cause gonad migration defects that prohibit staging animals precisely, animal age was determined

by the number of hours raised on RNAi post-arrest, and worms were confirmed to be at least post-mid-L2 by confirming *arIs51* expression in the AC. In addition, hermaphrodites were scored for appropriate hDTC defects caused by successful RNAi treatment: migration defects caused by *hlh-12* RNAi against animals with WT *lin-32*, and loss of migration and *lag-2p::rfp* expression caused by *hlh-12* RNAi against animals with *lin-32(0)* (see Figure 2.2). Due to the lack of migration in *hlh-3(0)* LCs, *lag-2p::rfp* expression was used to identify the proximal regulatory cell in the male gonad. *cdh-3p::gfp* was scored as "ON" if GFP expression was visible in the RFP+ cell at 250ms exposure, at which exposure expression can be seen in 100% of ACs beginning in late L2.

Pro-AC→other regulatory cell marker and morphology scoring

For scoring both marker expression and morphology changes, worms were evaluated at two stages. "Late L2" is equivalent to 26-28 hours post egg prep, and "late L3" is equivalent to 32-34 hours post egg prep at 25°C. Markers were considered "ON" at the following exposures:

qIs90: 800 ms, left scope (63x) or 50% power and 120ms, confocal (63x).
arIs51: 250ms, left and right scopes (63x).
hlh-12(1080)p arrays: 250ms, right scope (63x).

Morphologies were scored on the confocal (pro-AC→hDTC) or the left scope (pro-AC→LC). For pro-AC→hTDTC, morphology was considered "hDTC-like" if the cells featured flat cell bodies and long projections towards the germline. Morphology was scored using fluorescence from the *prox::lin-32::sl2::mcherry* transgene, imaged at 40% and 200ms. For pro-AC→LC, morphology was considered "LC-like" if the nucleus was round and featured a small nucleolus and stippling as seen in WT LCs, scored using Nomarski.

bHLH protein stability scoring

HLH-3::GFP and HLH-12::GFP array lines were scored using the right scope. Worms were synchronized using egg preps and grown at 25°C, and scored after 30 hours. GFP expression was considered “ON” at 250ms.

Table 3.1. Key resources table

Reagent	Source
GS8958: *arIs51; arIs222 him-5(e1490)*	This paper
GS8957: *hlh-3(tm1688); arIs51; arIs222 him-5(e1490)*	This paper
GS8959: *arIs51; arIs222 him-5(e1490); lin-32(tm1446)*	This paper
GS8991: *hlh-3(tm1688); arIs51; arIs222 him-5(e1490); lin-32(tm1446)*	This paper
OS8909: *nsIs497; unc-119(ed3); him-5(e1490)*	Shai Shaham, pers. comm
GS9309: *arSi45; qIs90; arTi148; arIs222 him-5(e1490)*	This paper
GS9148: *arSi45; arTi148; arIs222 him-5(e1490)*	This paper
GS9546: *arSi45; arTi148; arIs222 him-5(e1490); arEx2546[hlh-12(1080)p::gfp]*	This paper
GS9167: *arSi45; arTi275; arTi148; arIs222 him-5(e1490)*	This paper
GS9547: *arSi45; arTi275; arTi148; arIs222 him-5(e1490), arEx2547[hlh-12(1080)p::gfp]*	This paper
GS9149: *arSi45; arTi275; arIs222 him-5*	This paper
GS9095: *arTi276; arIs51*	This paper
GS9109: *arIs51; arTi148*	This paper
GS8952: *arTi275*	This paper
GS8518: *arTi148; ccIs4443*	Sallee et al., 2017
GS9467: *arIs222 him-5(e1490); lin-32(ar642)*	This paper
GS9474: *hlh-2(ar623); arTi112 him-5(e1490)*	This paper
GS9475: *arTi112 him-5(e1490); lin-32(ar642)*	This paper
GS9476: *arTi112 him-5(e1490); lin-32(ar633)*	This paper
GS9477: *hlh-3(vlc28); arTi112 him-5(e1490)*	This paper
GS9478: *hlh-12(ar643); arTi112 him-5(e1490)*	This paper
GS9479: *arTi112 him-5(e1490); hlh-8(ar644)*	This paper
GS3440: *pha-1(e2123)*	Karp & Greenwald, 2003
GS9548: *pha-1(e2123)*; *arEx2548[hlh-2prox::gfp::hlh-3]*	This paper
GS9549: *pha-1(e2123)*; *arEx2549[hlh-2prox::gfp::hlh-3]*	This paper
GS9550: *pha-1(e2123); arEx2550[hlh-2prox::gfp::hlh-3*	This paper
GS9551: *pha-1(e2123); arEx2551[hlh-2prox::gfp::hlh-12]*	This paper
GS9552: *pha-1(e2123); arEx2552[hlh-2prox::gfp::hlh-12]*	This paper

Results and Discussion

Experimental strategy: ectopic expression and stability of bHLH proteins in cells with AC potential

In this chapter, all transgenes were generated as single-copy insertion transgenes using either a miniMos transposon backbone or via CRISPR into a designated landing site on LGI (ref) and in the form of an artificial bicistronic operon as *hlh-2*prox*::bhlh::sl2acc::FP*, where "*sl2acc*" represents the intergenic region trans-splicing acceptor region from CEOPX036 and FP represents a fluorescent protein, either mCherry or 2xtagBFP2 (Figure 3.1A). The single copy insertion transgenes mitigated against overexpression, although the expression level driven by *hlh-2*prox appears to be lower than the expression of endogenous HLH-2 and the expression level of endogenous *bhlh* genes described below. The bi-cistronic design enabled me to visualize transgene expression while allowing for the bHLH proteins themselves to remain untagged, ensuring their proper function.

In WT animals, four cells are initially born with the potential to become the AC; over time, the two outer cells (β cells) specify as VUs, while the two inner cells (the α cells) undergo the LIN-12/Notch-mediated AC/VU decision (see Chapter 1). The four cells that have the potential to be the AC are termed "pro-ACs" by analogy with "proneural" (Karp and Greenwald 2004. While all four pro-ACs are born expressing HLH-2, protein expression is soon restricted to the α cells as they undergo the AC/VU decision, and then to the AC after AC specification (Attner et al., 2019). HLH-2 stability is restricted to the AC via proteasome-dependent degradation of HLH-2 dimers in the presumptive VUs (M. D. Sallee & Greenwald, 2015). Maria Sallee also found that some ectopically-expressed Class II proteins are degraded in the presumptive VUs, like HLH-2, while others are stable: for example, when LIN-32 is expressed using *hlh-2prox* the protein is only stable in the AC, while HLH-8 expressed in the same way is stable in the AC and VUs (M. D. Sallee &

Greenwald, 2015). Importantly, mutation of a residue required for dimerization of LIN-32 with HLH-2 prevented its downregulation outside the AC, suggesting that LIN-32 and potentially other Class II bHLH proteins can also be degraded in a dimerization-dependent manner in this context (M. D. Sallee & Greenwald, 2015).

This difference in stability can explain an interesting phenotype I saw with my pro-AC→mDTC transformation (see Figure 2.4B). Addition of HLH-8 alone results in expression of the HLH-8:HLH-2 heterodimer target *arg-1p::gfp* in a single cell with WT AC morphology, thus presumed to be the AC; the lack of *arg-1p::gfp* outside the AC suggested that HLH-2 was not stabilized in those cells by the addition of HLH-8. Addition of HLH-8+LIN-32 resulted in all four cells with AC potential expressing *arg-1p::gfp* and adopting mDTC-like morphology. This latter result can only be explained by stability of HLH2, HLH-8, and LIN-32 in all four cells, as we presume their activity leads directly to the transformation; as HLH-2 and LIN-32 are not stable outside the AC in WT animals, this means that they must be somehow stabilized in this background. One explanation is that HLH-8, which is stable in all four cells, is somehow able to stabilize HLH-2 and LIN-32 when combined with LIN-32, possibly by activating targets that block the dimerization-dependent degradation mechanism discovered by Maria Sallee. Sallee and Greenwald (2015) evaluated the stabilization of LIN-32 and HLH-8 in the AC and VUs. I found that GFP-HLH-3 was stable in only one cell, while GFP-HLH-12 was stable in up to four cells (Figure 3.2B). This raises the possibility that HLH-12:HLH-2 heterodimers may also be able to somehow stabilize HLH-2 and the other bHLH code genes when co-expressed with other members of the code, as HLH-8 appears to be doing in the pro-AC→mDTC transformation context, and thus the possibility that I may see more than one cell transformed with the addition of the hDTC or LC code to the pro-ACs.

I also assessed the individual *hlh-2prox::bhlh* transgenes for any changes in marker expression or AC function, which could indicate a change in the fate of cells with AC potential even if they do not possess the full bHLH complement of other regulatory cell types. Addition of the single bHLH code genes did not result in loss of the AC marker *cdh-3p::gfp* or cause any AC abnormalities (see Figure S3.1A, B, and Figure 4.4 for HLH-12 data).

pro-AC→hDTC transformation by expression of the hDTC code genes HLH-12 and LIN-32

I used *hlh-2prox* to express HLH-12 and LIN-32, the bHLH code for the hDTCs, in pro-ACs using the strategy outlined above, and assessed transformation by examining marker expression and morphological changes. As we do not have a single marker that distinguishes hDTCs from other gonadal regulatory cell types, I scored multiple markers to confirm that I was achieving transformation into an hDTC (Figure S3.1C). I inferred transformation to hDTC identity from expression of the reporter *qIs90*[*ceh-22b::venus*], which marks hDTCs and mDTCs, and *hlh-12(1080p)::gfp*, which marks the hDTCs and LC (Figure S3.1C, Lam et al., 2006; Tamai & Nishiwaki, 2007). While I note that we believe *hlh-12(1080)p::gfp* expression might be turned on from an autoregulatory loop driven by addition of HLH-12 alone (see below), this combination of markers still suggested a pro-AC→hDTC transformation.

In the L2 stage, when the AC/VU decision occurs, *ceh-22bp::venus* expression was observed in about 50% of hermaphrodites, most commonly in two adjacent proximal gonad cells (Figure 3.2A, top panel). This pattern suggests that the two α cells, which retain pro-AC potential and *hlh-2*prox expression for a longer time than the two β cells (M. D. Sallee & Greenwald, 2015), were transformed into hDTC-like cells. As *ceh-22bp::venus* expresses in both hDTCs and mDTCs, I confirmed transformation to hDTC-like fate with the marker *hlh-12(1080)p::gfp* (Figure 3.2A,

top; Figure S3.1C). I also saw morphological changes consistent with hDTC reprogramming: namely, flat cell bodies with long projections reaching towards the germline (Figure 3.2B). In addition, there was a significant correlation between *ceh-22bp::venus* marker expression and hDTC-like morphology, supporting the inference that these cells were transformed (Figure 3.2C).

As with the pro-AC→mDTC transformation effected by co-expression of HLH-8+LIN-32, transformation to hDTC fate caused by addition of HLH-12+LIN-32 is transient. By late L3, the proportion of individuals displaying *ceh-22bp::venus* marker expression in the proximal gonad is lessened (Figure 3.3A, bottom), and proximal gonadal cells do not display hDTC-like projections, although they do not look like normal ACs (Figure 3.3B). In contrast, when both LIN-32 and HLH-12 are expressed, the *hlh-12(1080)p::gfp* marker expression does not become apparent in the proximal gonad until the L3 stage (Figure 3.3B). Normally, this marker is expressed in specified hDTCs (and the LC), and endogenous HLH-12 is normally visible only after the hDTC is committed (Figure 3.4). Loss-of-function analysis is the principal basis for inferring that *hlh-12* contributes to hDTC specification: concomitant loss of *hlh-12* and *lin-32* led to apparent hDTC specification failure in ~50% of animals, as determined by both *lag-2p::rfp* and *hlh-2(fosmid)::gfp* expression (see Chapter 2, Figure 2.2D). Expression of *hlh-12* alone does not cause a cell fate transformation, although it does activate a program that allows the AC to move (Chapter 4), and *hlh-2*prox expresses through L3 (M. D. Sallee & Greenwald, 2015). These results would be consistent with positive autoregulation of *hlh-12* transcription initiated by the *hlh-2*prox*::hlh-12* transgene; expression of the endogenously-tagged *gfp::hlh-12* allele in the +HLH-12 AC would further support this hypothesis.

Addition of the full hDTC code to the pro-ACs appears to result in transformation in accordance with the code, based on both morphology and marker expression; the fact that marker

expression and morphological changes coincide makes a particularly strong case for transformation (Figure 3.3C). However, even in L2 I only see transformation in approximately 50% of the animals, and mostly commonly in only 2 cells despite the expected stability of HLH-12 in up to four cells (Figure 3.1B). This is in direct contrast to the pro-AC→mDTC context described earlier where I saw transformation in ~100% of animals and commonly in four cells, consistent with the stability expected of HLH-8 (Chapter 2). One explanation for this disparity is the fact that I drove HLH-8 in the pro-AC→mDTC context from a multi-copy extrachromosomal array, whereas here HLH-12 is driven from a single-copy integrated transgene: as *hlh-2prox* expression is weaker in the βs than the alphas, higher expression overall from the multi-copy +HLH-8 array was likely sufficient for transformation while lower expression from the single-copy +HLH-12 transgene was not. However, the transience of the transformation even in the pro-AC→mDTC context suggests the existence of barriers that are preventing the permanent reprogramming of the pro-ACs, which I will discuss in Chapter 5.

Late onset of LC-like features suggests transformation of the committed AC by expression of the LC code HLH-3+HLH-12+LIN-32, and possibly by subsets of this code

I used *hlh-2prox* to express HLH-3, HLH-12, and LIN-32, the bHLH code for the hDTCs, in pro-ACs. As with the hDTC, we had no marker that expressed only in the LC in the somatic gonad. The most common marker used to identify the LC is *lag-2p,* which also expresses in the AC (Chesney et al., 2009; Kato & Sternberg, 2009; Schwarz et al., 2012); I also evaluated *nhr-67(2kb)p::mcherry* (Shai Shaham, personal communication), and found that it too was expressed in the AC, precluding me from using it for our purposes. I therefore used *hlh-12(1080)p,* which is normally expressed in hDTCs and LCs, combined with morphological criteria to evaluate

transformation into LCs. The important caveat here is that we expect *hlh-12(1080)p* to have autoregulation, making its expression evidence of HLH-12 activity and not a fate transformation *per se;* however, in the absence of more LC-specific markers, I used it to determine if the cells were at least adopting some character consistent with LC fate.

Surprisingly, I did not observe any expression of *hlh-12(1080)p::gfp* expression or morphological alteration in the L2 stage (Figure 3.3A). However, by the late L3 stage, I found the marker was expressed in a single cell in the central, proximal region of the gonad in almost all hermaphrodites (Figure 3.3A), suggesting that the AC itself was taking on LC character. Furthermore, I was able to effect morphological transformation in these animals, with 8/20 L3 hermaphrodites having a single proximal gonad cell with LC-like morphology as assessed by Nomarski (Figure 3.3B). The fact that I see morphology changes in only a single cell suggests that in this context the AC itself is transformed into an LC, in contrast to the pro-AC→DTC contexts where the transformation happened before the cells acquired AC fate.

In addition, I saw LC-like morphology at around 10% penetrance in animals with only a subset of the LC code: either HLH-3+HLH-12, or HLH-12+LIN-32, again in only a single cell (Figure 3.3C). Importantly, I have not assessed *hlh-12(1080)p::gfp* in these backgrounds, and as we expect that *hlh-12(1080)p::gfp* may only require HLH-12 to turn on its expression in these backgrounds would not be conclusive anyway. However, based on the changes in morphology, these transformations suggest that the code heterodimers have some functional redundancy for LC specification, consistent with how we expect they are behaving in hDTC specification. While the presence of HLH-12 is common to both subsets, the morphology of +HLH-12 cells does not resemble either a WT LC or the cells transformed by these subsets, suggesting that HLH-12 must be combined with another LC bHLH code gene in order for the transformation to happen (Figure

3C). In addition, HLH-12+LIN-32 is the full hDTC code, which raises the question of whether these cells had transformed initially into hDTCs and from there into LCs, or whether these LC-like cells come from the subset of animals where I never see any transformation into hDTCs (Figure 3.2).

As with the pro-AC→DTC transformations, I find that the AC→LC transformation is transient. However, in this instance, the timing appears to be opposite: instead of being strong in late L2 and lessening by late L3, the morphological changes in the AC→LC reprogramming event do not occur before late L3 (Figure 3.3B). This is also true for the low penetrance of LC-like morphologies seen in animals with a subset of the LC (Figure 3.3C, data not shown). This change in timing suggests that there are fundamental differences in the mechanisms by which bHLH proteins promote LC fate transformation and transformation into either of the DTC types, and that transformation into an LC is delayed, takes longer than transformation into a DTC, or both.

My analysis of the AC→LC transformation and potential redundancy in the LC code was hampered by my lack of LC-specific markers. Unfortunately, I have found that markers expressed in the LC tend to be expressed as well in either the hDTC or the AC, likely due to the overlapping leader functions or proximal origin of the cell types (see Figure S3.1C for marker expression). A transcriptomics study of L4 LCs revealed several candidates for LC-specific markers (Schwarz et al., 2012); however, I could not see expression of fosmids for two of these genes, *srsx-18* and *smc-4*, in the LC (data not shown). In addition, we do not know if the LC-like cells I see in L4 resemble WT LCs shortly after their birth or WT L4 ACs; as this transcriptomics study looked only at L4 ACs, their candidate genes might not be expressed in my transformed cells. Future transcriptomics studies of earlier LCs might reveal markers that would be more useful for my purposes.

For a discussion of potential barriers to full reprogramming in this context, see Chapter 5.

Expression of endogenous bHLH genes: differences from fosmid reporters

I collaborated with Sarah Finkelstein in making three N-terminally tagged GFP knock-in alleles: GFP::LIN-32 [allele *lin-32(ar633)*], GFP::HLH-12 [allele *hlh-12(ar643)*], and GFP::HLH-8 [allele *hlh-8(ar644)*], and in addition obtained HLH-3::mNG [allele *hlh-3(vlc28)*] from Nuria Flames (Lloret-Fernández et al., 2018). All homozygotes and heterozygotes carrying these alleles are phenotypically wild-type, indicating that they do not significantly compromise gene function. Sarah Finkelstein analyzed the endogenous expression pattern of these proteins, and found some differences from the expression of the fosmid transgenes characterized previously (Figure 3.4, see Chapter 2).

One difference is in the expression of timing of LIN-32 and HLH-3. Fosmid reporters for both of these genes show their expression beginning in the specifying regulatory cells themselves, and not earlier (see Chapter 2, Figure S32.1). However, Sarah found that endogenously-tagged GFP::LIN-32 and HLH-3::mNG are visible earlier in the somatic gonad lineage, before the birth of the regulatory cells they specify: GFP::LIN-32 expression begins in the hDTC parents, and HLH-3::mNG expression begins in the grandparents of both potential LC lineages (Figure 3.4). This observation suggests that the inputs into bHLH code gene expression may not be in the somatic gonadal regulatory cells themselves, but rather their parent or grandparent cell.

The second difference is that, while in most cases the endogenous CRISPR tags of bHLH factors were seen in the regulatory cells predicted by the earlier fosmid expression analysis, neither Sarah nor I could not detect GFP::LIN-32 expression in the LC or its parent lineages before, during, or after the time of somatic gonad primordium formation (Figure 3.4). As the *lin-32* fosmid analyzed in Chapter 2 has a C-terminal tag (Figure S3.2A), Sarah then examined a C-terminally-tagged LIN-32::GFP [allele *lin-32(ar642)*], and similarly was not able to detect its expression in

the LC or its parent lineages even at the timepoint in L3 with clear LIN-32::GFP expression from the fosmid (see Figure S3.2B). In the next section, I describe genetic evidence that *lin-32* is relevant to LC specification.

Genetic evidence that *lin-32* activity influences LC development

There are several explanations for the apparent lack of GFP::LIN-32 expression in the LC according to our endogenously-tagged alleles. Firstly, there are differences between the *lin-32::gfp* allele and the fosmid, including sequence, transgene type, genomic locus, and copy number (Figure S3.2A); in particular, the single-copy endogenously-tagged lines might be expressing GFP::LIN-32 at levels too low to see on our scope. However, there is also the possibility that the fosmid might be expressing erroneously in the LC due to its higher copy number, potentially missing regulatory elements, insertion site, or the structure of the tag itself (for example, our tag contains a TEV site, while the fosmid tag has a 2xT1 element but no TEV site), and that our endogenously-tagged lines thus reflect a true lack of LIN-32 expression in the LC.

As an alternative approach to determine if LIN-32 is required for specification of the LC, I turned to genetics. As shown in Chapter 2, *lin-32(0)* animals have no defects in leader function in either males or hermaphrodites. The extension of the gonad posteriorly suggests that LIN-32 is not required for LC specification, and we could not assess the connection to the cloaca by the ability to sire progeny because males do not mate, possibly due to their abnormal tails (Portman & Emmons, 2000b). However, as discussed in Chapter 2, loss of both HLH-12 and LIN-32 leads to apparent loss of hDTC specification as assessed by both outgrowth failure and lack of HLH-2 and *lag-2* reporters, raising the possibility that there could be a similar synthetic requirement for LC specification.

In principle, functional redundancy with *lin-32* could be provided by *hlh-12*, *hlh-3*, or both. *hlh-12(0)* alone has abnormal leader function, but no specification defect of either hDTCs or LCs. *hlh-3(0)* is completely defective in LC leader function and the male proximal gonad regulatory cell expresses *hlh-2prox*, a marker of AC potential (Figure 2.4A), but not *cdh-3p::gfp*, a marker of differentiated AC fate (see photomicrographs in Figure 3.5), suggesting that loss of *hlh-3* may give the prospective LC the potential to adopt an AC fate but does not cause a full transformation into an AC. This raises the possibility that removal of further bHLH code genes from the LC, in particular *lin-32,* might further its conversion into a full AC.

To determine if there is a synthetic requirement for LIN-32 in LC specification, I assessed males for expression of the AC-specific marker *cdh-3p::gfp* in the proximal regulatory cell in different bHLH knockdown backgrounds (Figure 3.5, Figure S3.1C). I found that around 20% of *hlh-3(0); lin-32(0)* animals expressed *cdh-3p::gfp* in the male proximal regulatory cells, and that the phenotype became significantly more penetrant with added RNAi against *hlh-12* (Figure 3.5). I did not see *cdh-3p::gfp* expression in the male proximal gonad cell in the *hlh-3(0)* background with control lacZ RNAi, confirming our previous findings; I also saw no *cdh-3p::gfp* expression in *hlh-3(0); hlh-12(RNAi)* or *lin-32(0)*; *hlh-12(RNAi)* males; suggesting that the loss of both HLH-3 and LIN-32 is necessary to transform the erstwhile LC into an AC instead. As I have only done two RNAi trials to date, a third is necessary to confirm if the apparent difference between *hlh-3(0); lin-32(0)* and *hlh-3(0); lin-32(0); hlh-12*RNAi is significant, as it appears currently. However, while I cannot determine if this is a cell-autonomous effect with this approach, and thus if the LC itself expresses LIN-32, these preliminary results are consistent with a genetic requirement for LIN-32 in LC specification.

In WT hermaphrodites, *cdh-3p::gfp* expression begins in the AC shortly after its specification, and is clearly visible by late L2. Interestingly, expression in the *hlh-3(0); lin-32(0)* or *hlh-3(0); lin-32(0); hlh-12(RNAi)* LCs only became evident in mid-late L3, a full larval stage later than in their sister hermaphrodites (Figure 3.6A, hermaphrodite data not shown). This later expression pattern indicates that, although the LC takes on AC character, its initial specification has not been altered, further suggesting LIN-32 is not required in the somatic primordium for LC specification. This also forms an interesting parallel to the pro-AC→LC transformations discussed above, where transformation was only seen in late L3 as opposed to both pro-AC→DTC transformations seen clearly in L2, and suggests an inbuilt delay in switching between AC and LC fates, possibly due to the cells initially specifying as the endogenous cell type before the fate is altered.

Additional evidence against a role for *lin-32* in LC specification at the time of somatic primordium formation comes from the observation that *hlh-2(bx108); lin-32(0)* males have LCs (Figure 3.6B). *hlh-2(bx108)* is a hypomorphic allele that interferes with HLH-2 dimerization, and provides a sensitized background for loss of Class II gene activity (Portman and Emmons). For example, we performed an RNAi screen against bHLH genes in the *hlh-2(bx108)* background, and found that both *hlh-12(RNAi)* and *lin-32(RNAi)* showed phenotypes (see Chapter 2). The *hlh-2(bx108); lin-32(0)* LCs migrate, though they do show defective migration also seen in *hlh-2(bx108)* alone (data not shown); the fact that they do migrate overall suggests that *lin-32* is not required for LC specification in the somatic primordium, as if it were I would expect to see some failure to specify in this background. However, our results suggest that the loss of both *hlh-3* and *lin-32* can lead to the male proximal regulatory cell becoming a differentiated AC instead of an LC, albeit later in development than specification of the WT LC or AC.

In the pro-AC→other transformation contexts, I assessed transformation based on both marker expression and morphological changes. As the male proximal cells are displaced within the gonad both anteriorly and laterally in *hlh-3(0)* animals, likely due to the retained germline proliferation from the WT mDTCs (see Figure 2.3A), comparison of morphology to either a WT LC or AC is quite difficult. I found overall that the *cdh-3p* expressing cell was quite angular, more characteristic of an AC than the typically large and round LC (Figure 3.5), but due to the different placement and surrounding tissues I was unable to confidently assess whether the morphology was fully reminiscent of a WT AC.

In addition to morphology changes and marker expression, I would expect a WT AC to be able to induce VPC fate. The male hypodermal cells P3.p-P8.p have retained the competence to respond to EGFR signaling and form vulvas, and in several genetic backgrounds where otherwise male animals have ACs the males will form pseudovulvae from these hypodermal cells. For example, the forkhead box gene *fkh-6* is required for proper specification of the male proximal gonad (Chang et al., 2004; Sarah Finkelstein, personal communication). *fkh-6(0)* males have an AC instead of an LC ~97% of the time, as assessed by *cdh-3p* expression, and ~25% of *fkh-6(0)* males form pseudovulvae (Chang et al., 2004). In contrast, I did not see any pseudovulvae with my bHLH knockdowns (n>50 for each genotype and treatment condition).

There are a few different explanations for this lack of pseudovulvae. The first interpretation is that the "AC" in these animals does not secrete LIN-3/EGF to induce vulval fates in the male hypodermal cells, or that they do secrete LIN-3/EGF but are located too far away from the male hypodermal cells for the signal to induce their fate. However, another explanation for the lack of pseudovulvae is that the "ACs" specify too late. I find that the bHLH(-) cells only begin to express *cdh-3p* in L3 (Figure 3.6A). In WT males, the Pn.p cells either fuse with the hypodermis in late

L1 or late L3, except for those that go on to form the specialized male hook structures (Emmons, 2005). If the bHLH(-) cells are able to secrete LIN-3/EGF, they are likely doing it around the time that the hypodermal cells are fusing, meaning that the cells that would form a pseudovulva in this context are simply no longer there. An easy way to test this hypothesis would be to score expression of a *lin-3*/EGF marker such as the AC-specific promoter *lin-3*^{ACEL} (Deng et al., 2020; Hwang & Sternberg, 2004), possibly in conjunction with a hypodermal cell marker to judge when the male hypodermal cells have fused with regard to potential LIN-3 expression onset. Using this combination of markers would both further confirm that the bHLH(-) cells have acquired AC fate and support the hypothesis that these cells have indeed become functional ACs, but possibly at the wrong time or in the wrong place to induce pseudovulvae.

Conclusion

In Chapter 2, I effected pro-AC→mDTC transformation by adding the appropriate bHLH code genes to the pro-ACs. Here, I show that transformation to the other regulatory cell types is possible using this same method, though with some unexpected differences. The pro-AC→hDTC transformation is transient and most apparent in the L2 stage, consistent with my findings from the pro-AC→mDTC transformation. However, the pro-AC→LC transformation instead appears to be an AC→LC transformation, as only a single cell appears to take on LC characteristics, and at a much later stage in development. In addition, examining the expression of endogenously-tagged bHLH alleles raised questions about the role of *lin-32* that was only partially answered by further genetic analysis. These experiments when combined with the pro-AC→mDTC transformation shown in Chapter 2 suggest both that the bHLH code is capable of effecting transformation of the pro-ACs to all other regulatory cell types, and that there are some barriers preventing permanent

reprogramming of the pro-ACs; in addition, the bHLH genes may be functioning in a different manner to promote LC fate than for promoting DTC fate, a hypothesis supported by the new information about the endogenous expression patterns of bHLH genes in each cellular context. I will discuss these points further in Chapter 5.

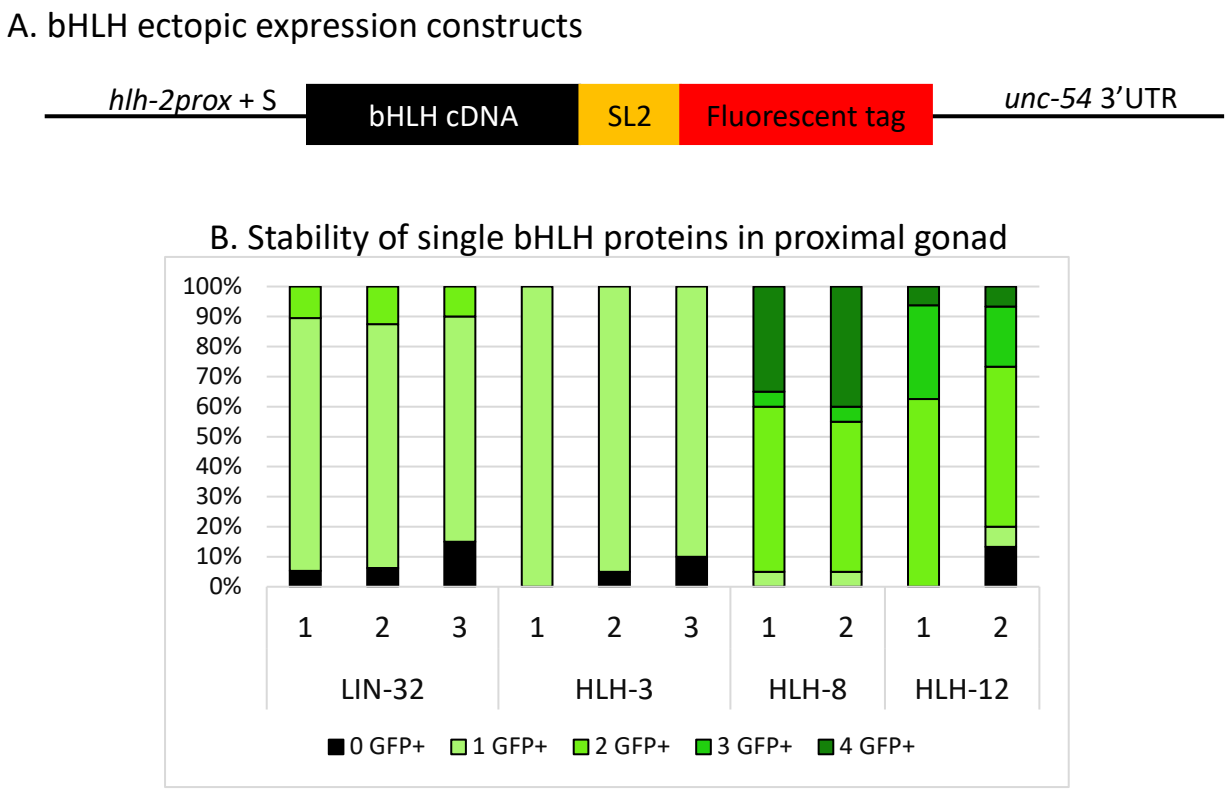


Figure 3.1. bHLH ectopic expression constructs and bHLH stability in the WT proximal gonad.

A. Cartoon schematic of bHLH ectopic expression constructs. All constructs were driven by *hlh-2prox* + synthetic intron "S" commonly used to boost expression level in *C. elegans* constructs (see Materials and Methods). The fluorescent tag for +LIN-32, +HLH-3, and +HLH-8 was mCherry, and for +HLH-12 was 2xtagBFP2. SL2 and *unc-54* 3'UTR sequences were the same for each construct. +LIN-32 and +HLH-3 constructs inserted into the genome using the miniMos transposon backbone, and +HLH-12 construct was inserted using the LGI CRISPR site backbone.

B. Stability of bHLH proteins in the proximal gonad of otherwise wild-type animals. Each bar represents a transgenic line carrying an individual extrachromosomal array. Note: LIN-32 and HLH-8 line data come from Maria Sallee (Sallee and Greenwald, 2015); lines for HLH-3 and HLH-12 were generated using her transgene conditions (see Materials and Methods for more details). n = 20-25 for each line scored (HLH-3 and HLH-12); n = 20 for each line scored (LIN-32 and HLH-8).

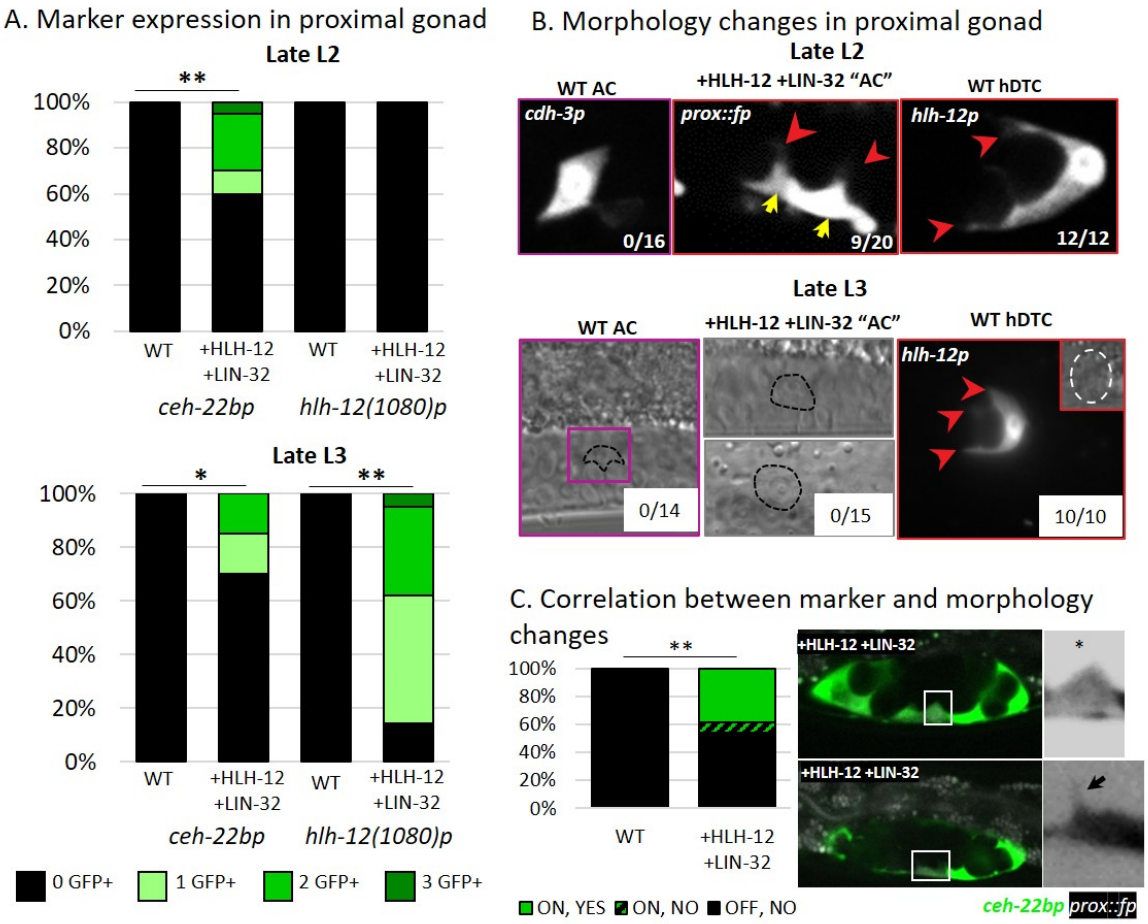


Figure 3.2. pro-AC→hDTC transformation.

A. Marker expression in pro-AC→hDTC animals in late L2 (top) and late L3 (bottom). n = 20-22 for each genotype. *p<0.05, **p<0.01, ***p<0.001, Fisher's Exact T Test.

B. Morphology changes in late L2 (top) and late L3 (bottom). Red arrows indicate projections; yellow arrows indicate cell bodies (top middle panel). n reflects animals showing projections. Top and bottom right panels are orthogonal projections. Top middle shows 2 of 3 transformed cells in this animal displaying germline projections. Bottom middle panels show range of proximal gonad cell morphologies in pro-AC→hDTC animals at this stage.

C. Correlation between marker expression and morphology changes. Marker used is *ceh-22bp::venus*. Graph legend refers to marker expression ("off" or "on") and morphological changes ("yes" or "no"). ** p<0.01, Spearman Correlation Test. Right photomicrographs are orthogonal projections showing representative pro-AC→hDTC cells; white boxes indicate area shown in black and white images to the right. Colored images show expression of both *ceh-22pb::venus* (green) and BFP from the +HLH-12 construct (white). For black and white images, both fluorophores are shown in black. Top: animal with two cells expressing *ceh-22bp::venus*, but no change in morphology. Bottom: animal with three cells expressing *ceh-22bp::venus* and a correlated change in morphology. White box outlines one expressing cell; black arrowhead points to projection; asterisk indicates lack of projections. n = 20-26 for each genotype.

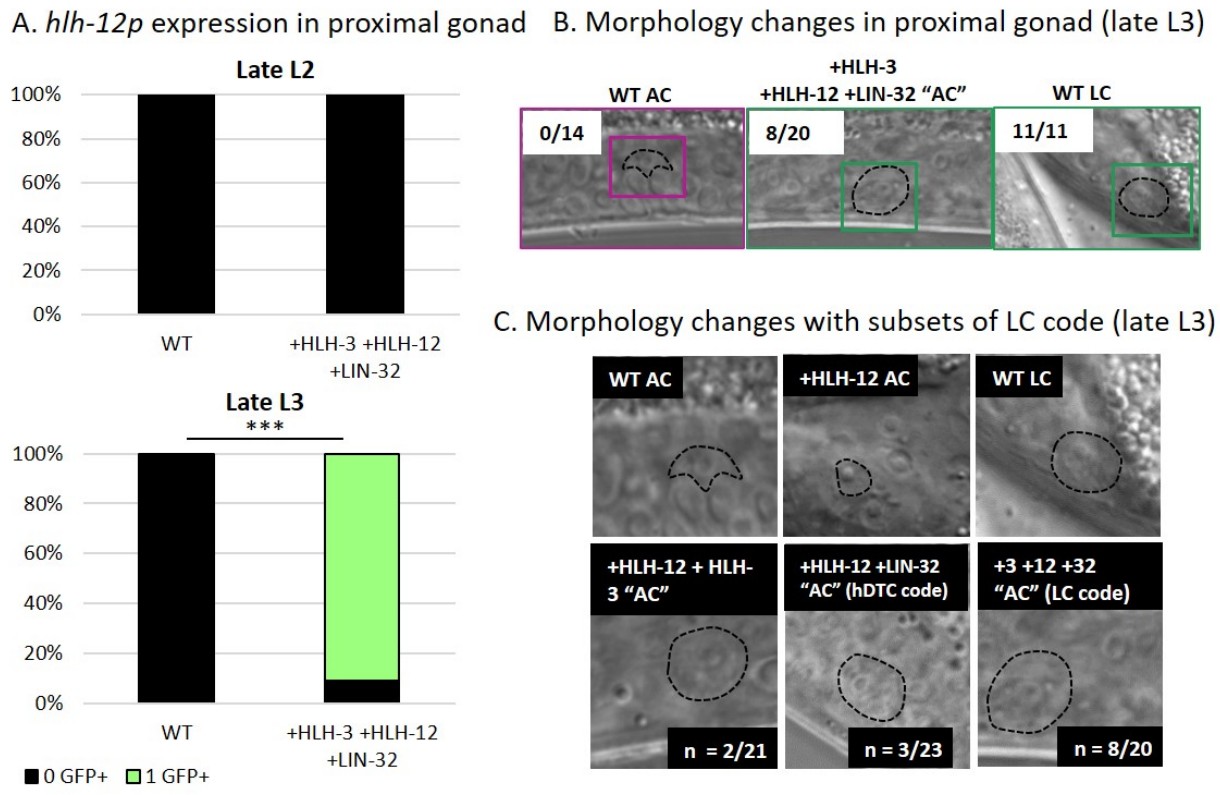


Figure 3.3. AC→LC transformation.

A. Marker expression in late L2 (top) and late L3 (bottom). ***p<0.001, Fisher's Exact T test. n = 20-22 for each genotype.

B. Morphology changes in proximal gonad, shown in late L3. n refers to animals with WT LC-like morphology; colored box and black dotted outlines indicate proximal gonad cells.

C. Morphology changes seen in hermaphrodites in which the AC ectopically expresses subsets of the LC code bHLH genes. Dotted black lines indicate proximal gonad cell. n in bottom panel refers to animals with LC-like morphology.

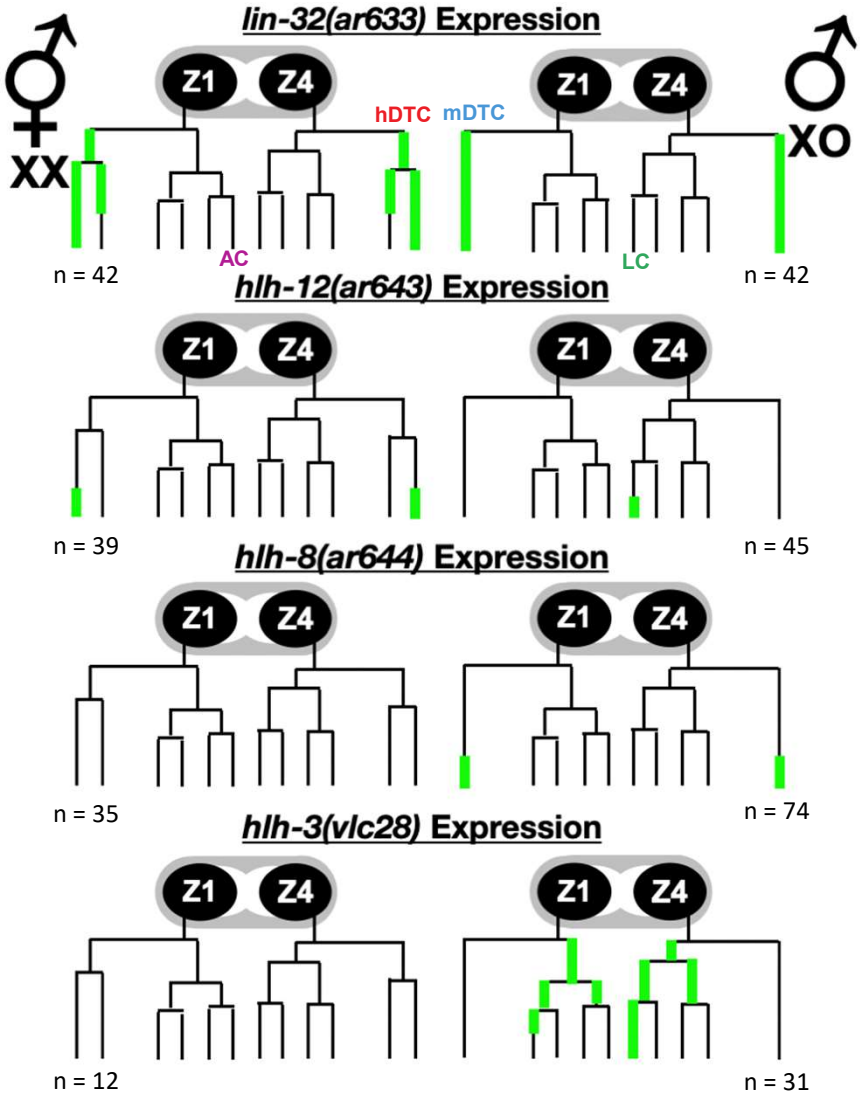


Figure 3.4. Expression of endogenous bHLH genes encoding fluorescently-tagged proteins. Green lines represent approximate onset of expression in each lineage; regulatory cell lineages are labelled in *lin-32* panel only. Note that GFP::HLH-12 expression begins in both the hDTCs and after LC after their fate is committed; schematics assume that Z4.aaa becomes the LC. n indicated on figure represents animals scored from 4-cell stage (Z1 and Z4 daughters) through somatic primordium formation. Adapted from Sarah Finkelstein.

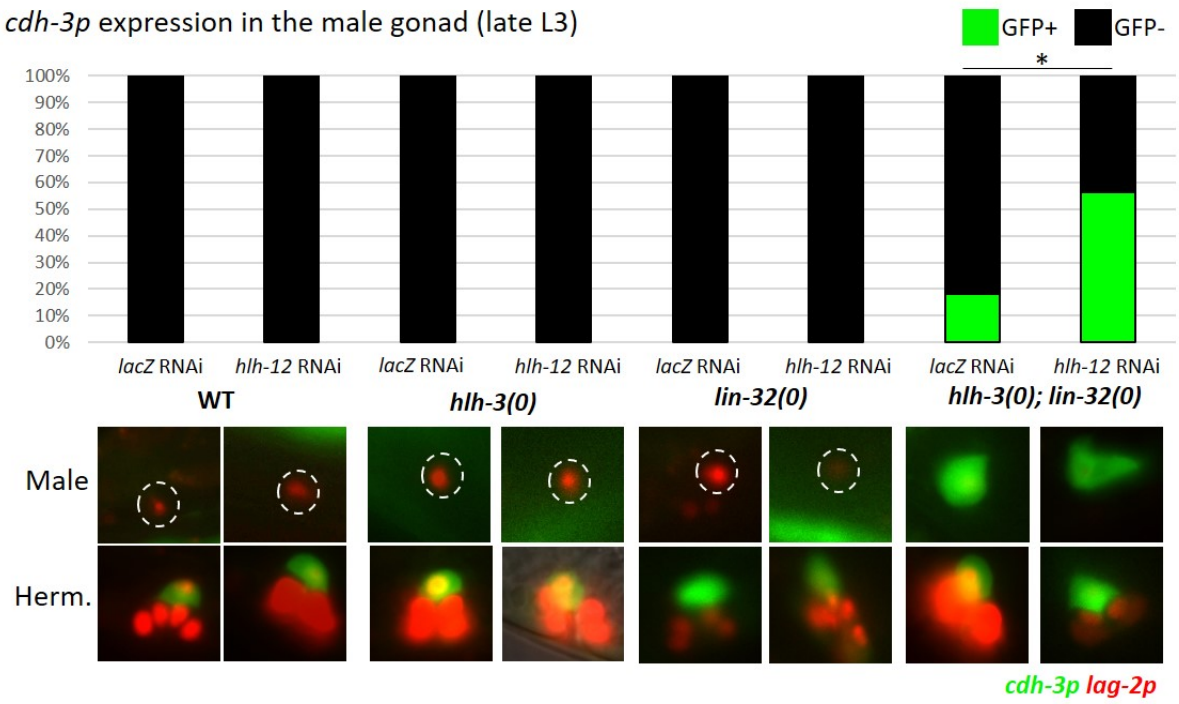


Figure 3.5. *cdh-3p::gfp* expression in the male proximal gonad. Top: graph represents average of two RNAi trials. Genotypes and treatments are indicated; n = 20-22 for each condition. * $p<0.05$ Fisher's exact T-test. Bottom: representative images of male proximal gonad cells and female ACs from each genotype. *lag-2p::rfp* cells in hermaphrodites are invaginating vulval cells; ACs are indicated by combined *cdh-3p::gfp* and *lag-2p::rfp* expression. Male proximal gonad cells are indicated by *lag-2p::rfp* expression (white outlines) in WT, *hlh-3(0),* or *lin-32(0)* backgrounds under both treatment conditions, or by *cdh-3p::gfp* expression in *hlh-3(0); lin-32(0)* animals with either *lacZ* or *hlh-12* RNAi.

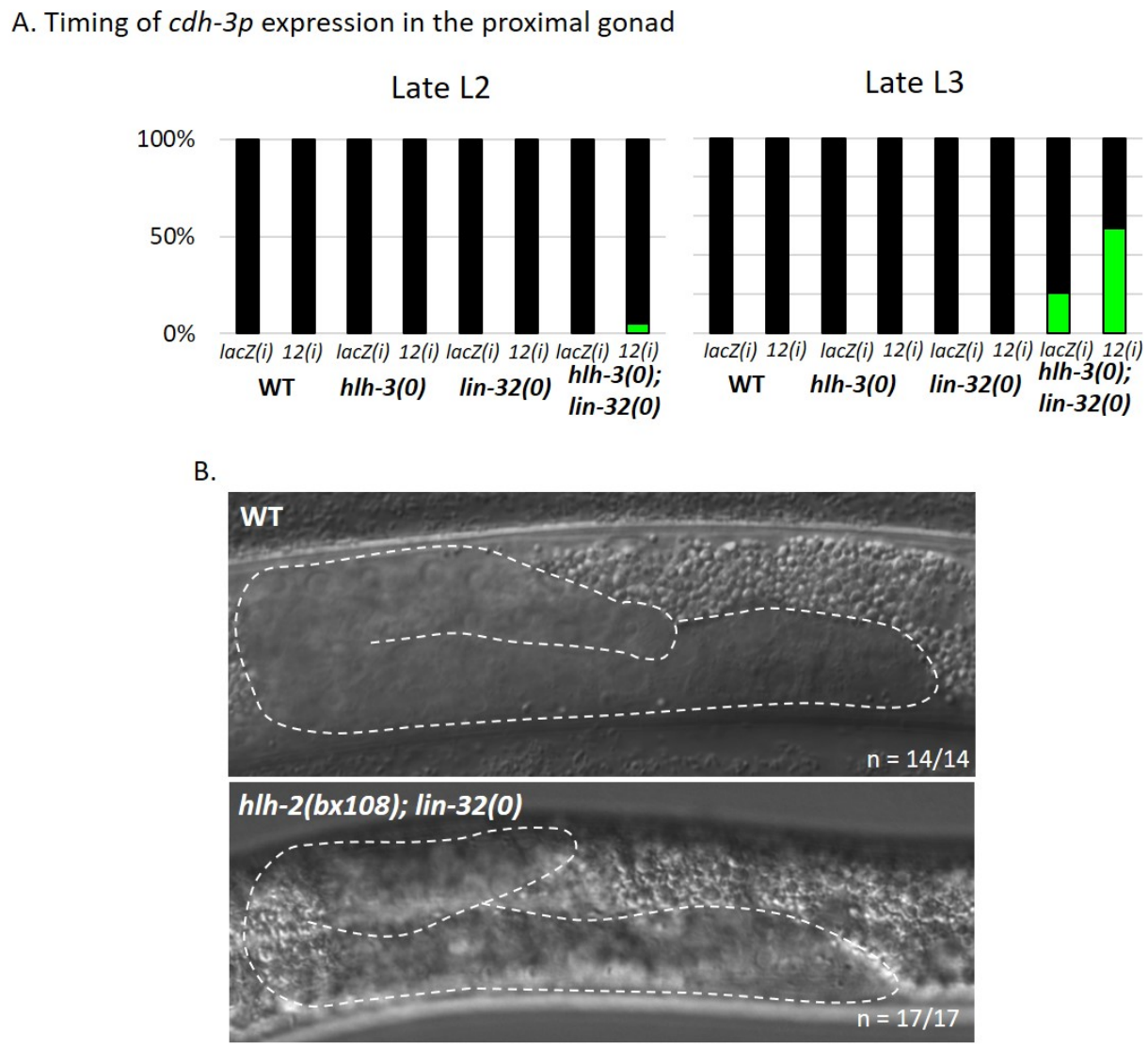


Figure 3.6. *lin-32* acts in LC specification after the somatic gonad primordium.

A. Timing of *cdh-3p::gfp* expression in the bHLH(-) male proximal gonad cells. Data shown are from one RNAi trial. (i) indicates RNAi treatment against either *lacZ* or *hlh-12*.

B. *hlh-2(bx108); lin-32(0)* LCs are still specified. n = animals showing extension of gonad arm. Top: WT L3 male gonad, showing proper extension and reflexion. Bottom: *hlh-2(bx108); lin-32(0)* male gonad, also in mid L3. Extension is abnormal due to the presence of *hlh-2(bx108)*, but gonad arms extend in all cases.

Supplemental Information

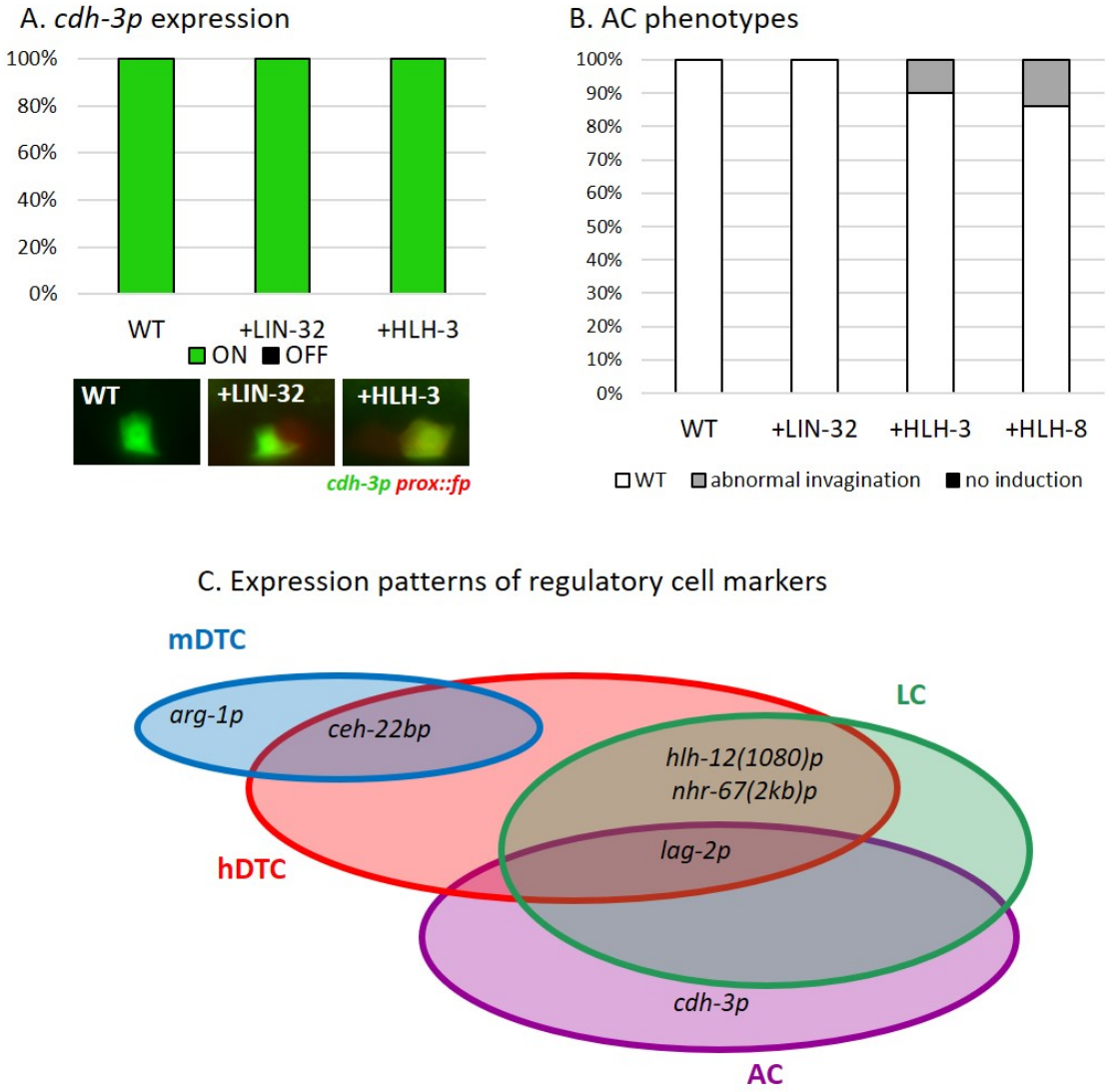


Figure S3.1. Single bHLH ectopic expression phenotypes and marker expression.

A. Addition of LIN-32 and HLH-3 does not affect morphology or *cdh-3p::gfp* expression. n = 20 for each condition. Photomicrographs show expression *cdh-3p::gfp* in only one cell, presumed the AC because of its morphology.

B. AC phenotypes with addition of single ectopic expression constructs. "Abnormal invagination" indicates delayed invagination or invaginations with abnormal morphology. n = 20-24 in each condition.

C. Expression of somatic gonad markers. For *arg-1p::gfp,* see Chapter 2. *nhr-67(2kb)p::mcherry* expression was assessed, but as it expresses in both the LC and AC it was not used further in this thesis.

A. *lin-32* CRISPR and fosmid constructs

B. *lin-32* transgene expression in the LC

Stage	C-term CRISPR expression	Fosmid expression
Late L1-L2	0/30	8/40
L3	0/14	17/18

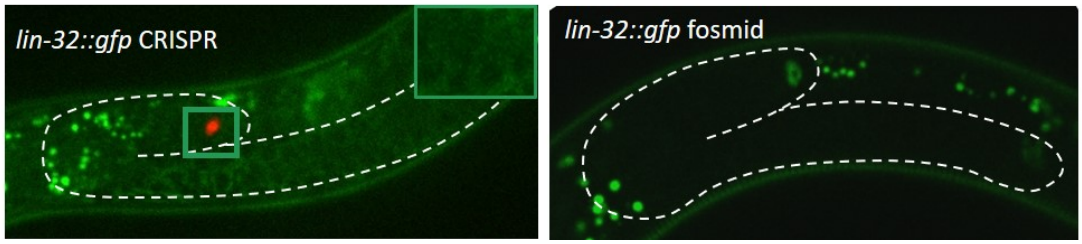


Figure S3.2. Comparison of *lin-32 (allele[lin-32::gfp])* and *labX[lin-32::gfp]* fosmid expression data.

A. Structure of *lin-32* endogenously-tagged alleles and *lin-32::gfp* fosmid. In addition, *lin-32* endogenously-tagged alleles are single-copy, while *lin-32::gfp* fosmid transgene is an integrated multi-copy array. All other N-terminally-tagged endogenous alleles have the same structure as the *lin-32* allele represented above. Note that the GFP sequences differ between the two CRISPR-generated transgenes and the fosmid transgene, represented by different shades of green in the schematics.

B. Comparison of expression between *lin-32::gfp* allele and *lin-32::gfp* fosmid. Photomicrographs show animals in L3: white dotted lines outline gonad, and green box outlines LC. Red in CRISPR-generated allele photomicrograph is from *lag-2p::rfp*. Inset (left) shows lack of GFP expression in the LC.

Chapter 4: Investigating the contribution of bHLH genes to migrating anchor cells in other nematode species

Abstract

All nematode gonads are patterned by the activity of cognate regulatory cells. However, in several species, different activities of the regulatory cells lead directly to gonads with strikingly different morphologies. As manipulating bHLH complements can change the fate and functions of *C. elegans* regulatory cells, I hypothesized that these different gonad morphologies might reflect different patterns of bHLH code gene expression in these other species. In particular, I found that addition of HLH-12 to the pro-ACs results in a cell which specifies AC fate but migrates via established hDTC/LC mechanisms, suggesting that expression of HLH-12 alone is sufficient for regulatory cell-type migration. As the distantly-related species *M. belari* has an AC which migrates, I hypothesized that expression of *Mbe-hlh-12* in this AC might explain its migratory ability. To determine if changes in the bHLH code might underlie the different morphology of the *M. belari* gonad, I first performed a phylogenetic analysis to identify bHLH code orthologs in other species. While I find conservation of most bHLH code genes, I establish that *hlh-12* appears to be novel to the genus *Caenorhabditis*. While this suggests that the *M. belari* AC cannot be migrating due to *hlh-12* ortholog expression, I instead predict that another bHLH gene may be controlling its migration, as *hlh-2* is highly conserved in all species studied and the *Ce-hlh-12* target genes have conserved *hlh-12*-type E boxes.

Introduction

bHLH factors in general are known to be highly conserved and are found in organisms ranging from yeast to humans (Massari & Murre, 2000). In yeast, bHLH factors regulate a number of developmental processes, from controlling metabolic processes to regulating DNA replication and morphogenesis (Chen & Lopes, 2010; Stoldt et al., 1997; Zhang et al., 2002). In metazoans, bHLH proteins are most often involved in developmental events, particularly including cell fate commitment and lineage specification; in humans, bHLH genes are notably involved in several different cancers such as B-cell lymphomas (Massari & Murre, 2000; Sun et al., 2007).

There have been several different attempts to further classify bHLH proteins by both sequence and function. Murre et al. (Murre et al., 1994) first sub-divided bHLH genes into six classes. Class I and Class II bHLH proteins have been discussed extensively in this work; of the rest, Class III includes myc-related proteins, Class IV those which interact with myc, Class V those without a DNA-binding domain, and Class VI those with a proline in their basic region such as *Hairy* and *Enhancer of split*. Their classification was later expanded to include Class VII, comprised solely of proteins containing both a bHLH domain and a PAS (Period-ARNT-Single-minded) domain (Massari & Murre, 2000). This classification system neatly separates bHLH proteins by their function, sequence and structural elements, and predicted dimerization partners and expression, but provides little information about any evolutionary relationship between the proteins.

A more evolution-based study was done by Atchley and Fitch (1997), who took a phylogenetic approach to a hand-picked selection of 122 amino acid sequences from species ranging from plants to humans, chosen specifically to maintain a relatively wide sequence divergence. They used only the bHLH domain in their study, as they found that the flanking

sequences were too divergent to provide meaningful analysis. This approach resulted in 27 different subfamilies, grouped into four clades that aligned with the functions which Massari et al. used to group their initial classes (Figure 4.1A). Atchley and Fitch's Group A corresponded exactly with the known Class I and Class II bHLH proteins included in their tree, and Group B with their Class III, IV, and VI proteins. In comparison, Groups C and D were much smaller, with the sole included Class V sequence included with Group D. Of these four groups, Group B is both the largest and contains proteins with the most divergent functions; Atchley and Fitch propose that Group B is the ancestral group of bHLH proteins, meaning that Class I and Class II bHLH proteins are both closely related to each other and likely represent a more evolved bHLH lineage.

This phylogenetic grouping of the bHLH genes was later refined by a second analysis done by Vervoort and Ledent (2002, Figure 4.1B). They included a total of 350 putative bHLH sequences from fly, mouse, worm, yeast, and plants, and found 44 bHLH gene families, 35 of which were common to all animals included. Vervoort and Ledent found that Atchley and Fitch Groups A and D were monophyletic, and Group C was in a paraphyletic clade with another novel group which they termed Group E. They also proposed the existence of a Group F, and found that Group B was paraphyletic as well, with the common ancestor of all Group B sequences being the common ancestor of the tree as a whole. As Atchley and Fitch's Group B was both proposed to be the ancestral group of bHLH genes and contained representatives with vastly different functions (see above), its paraphyletic identity according to Vervoot and Ledent makes relative sense. A more surprising finding of this study was the fact that Group C is paraphyletic as well. The Group C proteins in Atchley and Fitch's work all include a conserved bHLH-PAS domain, named after the common motif in *Drosophila* Period, human ARNT, and *Drosophila* Single-minded (Crews, 1998); all are Class VII according to Massari and Murre. Vervoot and Ledent suggest that the two

independent associations between the bHLH and PAS domains is a result of two separate domain shuffling instances, a method of bHLH gene evolution previously proposed by Morgenstern and Atchley (1999). However, in both phylogenetic studies the Class I and Class II bHLH proteins formed a monophyletic clade, further suggesting their close evolutionary relationship; this was underscored by the same finding by a third study using the full protein sequence from a range of organisms (Stevens, Roalson, & Skinner, 2008).

To date, there has been only one published phylogenetic analysis of the *C. elegans* bHLH genes. Grove et al. (2009) found that *C. elegans* has 18 Class II bHLH genes; of those, they classified *hlh-16, hlh-17, hlh-31,* and *hlh-32* as Class II based on similarity sequence to other Class II proteins despite a lack of convincing dimerization data. They find that Class II bHLH proteins are divided into two separate clades (Figure 4.2). The first, containing *hlh-3, hlh-4, hlh-6, hlh-14,* and *hnd-1*, forms one of three supergroups in the tree, along with a clade containing *hlh-1* and *hlh-2* and a third clade containing the rest of the predicted bHLH genes. Interestingly, they find that *hlh-12*, which they consider a Class II bHLH gene based on both sequence data and dimerization data, is most closely related to the Class VII bHLH-PAS gene *hif-1*, and is only distantly related to the other Class II bHLH genes. Again, this study used only the bHLH domains to form their tree, leaving the possibility for different clustering when the full bHLH sequences were taken into account.

Given the general high level of conservation seen in bHLH genes and the importance of their roles in gonadogenesis, we wondered if the bHLH code that we identified in *C. elegans* might also be functioning in the patterning of the gonads of other species. While the majority of gonads in nematode species follow a plan similar to that of *C. elegans*, with hermaphrodites or females having a bilaterally-symmetric, two-armed gonad with a vulva at the midbody and males having a

single-armed gonad terminating in the cloaca at the tail, there are some notable differences to this basic scheme. However, all nematode gonads are patterned by the same regulatory cells, and differences in number or function lead to the different gonad shapes (Félix & Sternberg, 1996, Figure 4.3). For example, the most basal *Caenorhabditis* species, *C. monodelphis*, has a single-armed gonad resulting from the death of one female DTC (fDTC) (Slos et al., 2017, Figure 4.3). *Oscheius guentheri*, a more distantly related species, has a two-armed asymmetric gonad which varies between individuals. The anterior fDTC always makes both turns and terminates dorsal to the vulva, as in *C. elegans*, but the anterior fDTC ranges in migration from terminating shortly post-vulva to making an almost-complete arm; while there is always some germline found in the posterior arm, even those of almost-anterior size remain sterile (Sudhaus & Hooper, 1994; Félix & Sternberg, 1996; Figure 4.3). Interestingly, these differences in fDTCs appear to go up the lineage: Z1, the progenitor of the anterior fDTC, is heavily biased towards dividing first in *O. guentheri*, but there is no such bias in the related didelphic species *Oscheius* PS1131 (M. a Félix & Sternberg, 1996). This suggests that there are differences in both lineage and A/P positioning between the two fDTCs, which may jointly contribute to their different levels of activity.

There is also evidence for differences in AC function between species. In *C. elegans*, the AC is required for all stages of vulval formation: it must first induce primary fate in the appropriate vulval precursor cell (VPC), followed by basement membrane invasion and coordination of further vulval development (Sternberg, 2005). Interestingly, the AC is not the sole source of inductive signal in many more basal species (M. A. Félix et al., 2000; M. A. Félix & Sternberg, 1997). For example, in *Panagrolaimus* species PS1732, the first inductive signal comes from the somatic gonad precursor cells Z1 and Z4 and specifies P5.p-P8.p as vulval; later, after is it born, the AC sends a second inductive signal to specify primary vs. secondary fate within the pre-specified

VPCs (M. A. Félix & Sternberg, 1997). This coordinated two-wave induction system is also present in several other nematode species outside *Caenorhabditis* (M. A. Félix et al., 2000; Kiontke et al., 2007a). Interestingly, however, some species seem to have dispensed with the AC's role in induction altogether, as induction in the *Mesorhabditis* group of species and in *Diplogastrellus gracilis* is gonad-independent (Kiontke et al., 2007; Figure 4.3). Kiontke et al. show that the source of the first inductive signal has changed at least four times over the course of evolution, and propose that heterochronic changes in other pathways required for induction, such as EGF receptor expression in the VPCs, might underlie these evolutionary shifts in AC function.

For this thesis, I was particularly interested in a group of species in which the AC appears to have gained the ability to migrate. In one particular species, *Mesorhabditis belari*, the vulva is located towards the posterior instead of at the midbody. In these animals, the AC specifies at the posterior instead of the center of the somatic gonad primordium, opposite the sole fDTC (Félix & Sternberg, 1996; Figure 4.3). As discussed above, vulva formation is gonad-independent in this species (Kiontke et al., 2007a; Sommer & Sternberg, 1994). Instead, the AC migrates posteriorly along the ventral side of the animal during L2 and L3, and connects the forming uterus to the vulval opening (Félix & Sternberg, 1996; Figure 4.3). Thus, in *M. belari* the female AC appears to act more like the *C. elegans* male LC: both cell types lead extension of the arm and connect to the gonad opening—the vulva in *M. belari* females, and the cloaca in *C. elegans* males—without inducing cell fate.

As both bHLH factors and regulatory cell identity in nematodes are so conserved, I hypothesized that the bHLH code might be regulating the specification and function of regulatory cells in other species, and, further, that changes in regulatory cell identity or function might be explained by differences in their bHLH complements. I was particularly interested in the migrating

and non-inducing *M. belari* AC, which functionally resembles the *C. elegans* LC. I first hypothesized that expression of *Mbe-hlh-3* in the AC might give it LC character, as I found previously that *hlh-3* is necessary for LC identity in *C. elegans* (see Figure 2.4). However, ectopic expression of *hlh-3* in the pro-ACs did not affect AC identity or cause any LC-like changes that we could see (see Figure S3.1), suggesting that it is not sufficient for LC identity. Instead, I hypothesized that expression of an *hlh-12* ortholog in the *M. belari* AC might be granting it migratory functions, as in *C. elegans hlh-12* promotes migration of both the hDTCs and LC in a cell-autonomous manner (Blelloch & Kimble, 1999; Meighan & Schwarzbauer, 2007; Tamai & Nishiwaki, 2007a).

As described below, I found that addition of CE-HLH-12 to the *C. elegans* pro-ACs resulted in cells that migrated in an hDTC/LC-like manner, but which retained their AC fate and most AC-specific functions, supporting the hypothesis that expression of *Mbe-hlh-12* in the *M. belari* AC enables its migration. Given the similarity of Class II genes, I used a phylogenetics-based approach to properly identify their orthologs. As expected, I found that bHLH genes in general were quite conserved across nematode species, and in particular *hlh-2* and four of the five bHLH code genes have orthologs in most or all of the species studied. The one exception is *hlh-12*, which I do not find outside of *Caenorhabditis* despite its location at a relatively conserved genomic locus; further investigation shows that it most likely arose near the base of the *Caenorhabditis* clade, in the species *Caenorhabditis bovis*. As both the *Ce-hlh-12* target genes and *hlh-12*-type E boxes in their 5' regulatory regions are conserved, I hypothesize that activity of a different Class II dimer might control the migration of regulatory cells in other species, including the *M. belari* AC.

Materials and methods

See Figure 4.7 for phylogenies of species used in trees, Table 4.1 for strains used in this study, Table 4.2 for RNAi clones used in this study, Table 4.3 for annotated transcripts, Table 4.4 for genome assemblies used in BLAST search, and Table S4.1 for BLAST search results.

C. elegans genetics

Strain names and full genotypes are listed in Table 1. For details on strain maintenance, see Chapter 2; strains were analyzed at 25ºC. The allele *fos-1(ar105)* contains an early termination codon, and is a likely null allele (Seydoux et al., 1993; Sherwood et al., 2005). The allele *ina-1(ar639[ina-1::gfp])* was generated using CRISPR/Cas9 as described below.

The following alleles and transgenes were used:

arIs51[cdh-3p::gfp] (Karp and Greenwald, 2003) is expressed in the AC after its specification in late L2 and through L3, and later in the AC and developing uterine seam cell (utse) in L4.

arIs222[lag-2p::2xlns-tagRFP] and *arIs131[lag-2p::2xnls-YFP]* (Sallee and Greenwald, 2015; Li & Greenwald, 2010) are expressed in all somatic gonad regulatory cells except for the mDTCs, as well as P6.p and its descendants. I will refer to *arIs222* as "*lag-2p::rfp"* and *arIs131* as "*lag-2p::yfp"* below.

qIs90[ceh-22bp::venus] (Lam et al., 2006) expresses in the hDTCs and their sisters, and the mDTCs.

qyIs176[zmp-1p::mcherry-moesinABD] (Schindler & Sherwood, 2011) contains the moesin actin-binding domain and localizes to the AC invasive membrane. It is driven by *zmp-1p*, which expresses in the AC beginning in L3, as well as the LC.

bmd138[gfp::fos-1] (Medwig-Kinney et al., 2020) is an endogenously-tagged allele generated through CRISPR, expressed in the hermaphrodite proximal gonad beginning in L3, and the LC.

ar639[ina-1::gfp] (this study) is a CRISPR allele expressed in the basement membrane of gonads of both sexes beginning in L1, the hDTC and LC beginning in L2, and the hermaphrodite proximal gonad beginning in late L3.

fos-1(ar105) contains an early termination, and is a likely null allele (Seydoux et al., 1993).

arSi45 is a single-copy transgene made using the LG1 integration site (Pani and Goldstein, 2018), which contains *hlh-2prox::hlh-12cDNA::sl2::2xtagBFP2*.

Constructs and generation of single-copy insertion transgenes

arSi45[hlh-2prox::hlh-12cDNA::sl2::2xtagBFP2]: *hlh-12* cDNA (genewiz.com) and *hlh-2prox* regulatory sequences amplified from pHL4 (Sallee et al., 2017), were combined using fusion PCR (Hobert, 2002), and inserted into SmaI/XbaI-digested pBS SK+ (Mayer, 1995) using Gibson assembly (Gibson et al., 2009) to create pHL48. The *unc-54 3'UTR* was amplified from pHL4; *2xtagBFP2* (hereafter "*BFP*") was amplified from pZW109 (Dickinson et al., 2015), with the addition of the first 15bp of the 2xGGGGS linker sequence plus an artificial ochre stop codon to allow for better amplification of the repetitive sequence. The two pieces were inserted into SmaI/XbaI-digested pBS SK+ using Gibson Assembly to create pHL50. The *sl2* trans-splicing sequence was amplified from pHL3 (Sallee et al., 2017) and fused together with *BFP::unc-54 3'UTR* amplified from pHL50 to create *sl2::BFP::unc-54 3'UTR*. *hlh-2prox::hlh-12cDNA* was amplified from pHL48. *hlh-2prox::hlh-12cDNA* and *sl2::BFP::unc-54 3'UTR* were inserted into SpeI/AvrII-digested pZW111 (Pani and Goldstein, 2018) using Gibson assembly to create pHL53.

The insert from pHL53 was inserted using CRISPR/Cas9 into a defined site on LGI using the method of Pani and Goldstein (2018). Hermaphrodites were injected with 55ng/µL pAP82, 10ng/µL pGH8, 2.5ng/µL pCFJ90, and 10ng/µL pHL53. Injected P0 were placed at 25°C and treated with 400µL 5mg/mL hygromycin three days after injection. Approximately 100 L1 from plates with homozygous "dark rollers" were heat-shocked at 34°C for four hours to excise the roller cassette, and lines were established from single non-rolling L4 progeny of heat-shocked animals.

ina-1(ar639[ina-1::gfp]. The homology repair template pHL57 was cloned into pDD282 digested with AvrII and SpeI using the methods described in Dickinson et al. (2015). pHL58 contained the sgRNA, and was constructed by digesting pJW1285 (Ward, 2014) with NdeI and SpeI. PCR fragments containing the missing portion of pJW1285, but with the desired *ina-1* sgRNA replacing the original *pha-1* sgRNA contained in pJW1285, were inserted into this digested backbone using Gibson assembly. This resulted in a new vector identical in sequence to pJW1285, with the exception of the desired *ina-1* sgRNA sequence in place of the original *pha-1* sgRNA. Microinjections were performed using the method described in Dickinson (2015); template and sgRNA constructs were each injected at 50ng/µl, and lines were isolated as described above.

Scoring cell fate transformation and function

All worms were synchronized for scoring following the methods described in Chapters 2 and 3, using a bleach/sodium hydroxide solution. Experiments were scored using widefield microscopy, with either a Zeiss Axio Imager D1 microscope (40x or 63x PlanNeofluar) and a Zeiss AxioCam MRm camera, or a Zeiss Axio Imager Z1 microscope with (40x or 63x PlanNeofluar

objectives) and a Hamamatsu Orca-ER camera. In addition, we used a spinning disk confocal microscope (Carl Zeiss) with a 63x PlanApochromate objective to better visualize cellular morphology.

Pro-AC→fDTC morphology and marker scoring.

Morphology: Worms carrying *arTi14[hlh-2prox::lin-32cDNA::sl2::mCherry]* and/or *arSi45[hlh-2prox::hlh-12cDNA::sl2::2xtagBFP2]* were evaluated at two stages. "Late L2" is equivalent to 26-28 hours post egg prep or 8-10 hours post-L1 arrest at 25ºC, and "late L3" is equivalent to 32-34 hours post egg prep or 22-25 hours post-L1 arrest at 25°C. Morphological changes were assessed using the confocal microscope, and cells were considered to have fDTC-like morphology if they displayed flat cell bodies and long projections towards the germline. Visualization of cell morphology was facilitated by fluorescence from the *hlh-2prox::lin-32::sl2::2xtagBFP2* transgene, imaged at 40% laser power and 800ms exposure time.

Markers: Markers were evaluated using widefield microscopy and considered "ON" at the following exposures: *qIs90*[*ceh-22b::venus*]: 800 ms (Axio Imager D1 microscope, 63x objective) or 50% laser power and 120ms, spinning disk confocal (63x). *arIs51[cdh-3::gfp]*: 250ms, both widefield microscopes (63x objective).

To examine the correlation between fDTC-like morphology and *ceh-22bp::venus* expression, we examined hermaphrodites carrying both *arTi148[hlh-2prox::lin-32cDNA::sl2::mCherry]* and *arSi45[hlh-2prox::hlh-12cDNA::sl2::2xtagBFP2]* using spinning disk confocal microscopy; to test statistical significance, we generated a 2x2 contingency table and used a Fisher's exact T test, with morphology scored as fDTC-like or WT and *ceh-22bp::venus* expression scored as ON or OFF under the conditions described above.

+HLH-12 AC scoring.

The ACs of hermaphrodites carrying *arSi45[hlh-2prox::hlh-12cDNA::sl2::2xtagBFP2]* were scored for features characteristic of the wild-type AC using markers and morphology indicated here.

Invadopodia: Invadopodia were defined by colocalization of membrane protrusions and mCherry-moeABD punctae. Worms were imaged using the spinning disk confocal microscope at 63x in late L3. For WT, *qyIs176[zmp-1p::mcherry-moeABD]* was imaged at 40% power and 800ms; for +HLH-12, it was imaged at 50% power and 1200ms. *arIs131[lag-2p::2xnls-YFP]* was imaged at 20% power and 500ms in both backgrounds.

Vertical displacement scoring: Vertical displacement was measured using ImageJ (fiji.sc).

Quantification of displacement distance (see Figure 4.5A, Figure S4.2B): Hermaphrodites at the P6.pxx stage of the vulval lineage were scored for the degree of displacement of the AC from the center of the P6.pxx nuclei, as shown in Figure 5A and Figure S2B. Migration distance is expressed as a ratio of the horizontal distance from the center of the AC nucleus to the outside edges of the P6.pxx nuclei, with a positive ratio indicating anterior placement and a negative ratio indicating posterior placement. We note that the WT AC is slightly displaced towards the anterior, with an average ratio of +0.27, which we refer to as its "normal position". All distances were measured images using FIJI, with a macro to create perpendicular lines intersecting at the midpoint of previously saved lines. Nuclei were marked with *lag-2p::rfp* or *lag-2p::yfp* expression, or determined by Nomarski, and midpoints of nuclei were determined by eye. For statistical analysis, animals were sorted into 3 bins depending on the location of the AC relative to the individual P6.pxx nuclei (see Fig S4.2B for details). Animals were sorted into bins using the absolute value

of their displacement ratio, meaning that animals with anterior and posterior displacement were combined into the same bin.

Marker expression: All markers except *ina-1(ar639[ina-1::gfp]* were scored on using widefield microscopy at 63x and scored as "ON" if fluorescence was visible at the following exposures: *arIs51[cdh-3::gfp]* (250ms), *qIs90*[*ceh-22b::venus*] (800 ms), *fos-1(bmd138)* (300 ms). Expression of *ina-1(ar639[ina-1::gfp])* was scored using confocal microcopy (30% power and 500ms, 63x).

Effect of loss of *fos-1* activity: *fos-1(ar105)* was maintained as a heterozygote balanced by *tmC16[unc-60(tmIs1210)]. tmC16* shows green fluorescence as a heterozygote and a homozygote and is Unc as a homozygote (Dejima et al., 2018). *fos-1(ar105)* homozygous worms were thus identified as non-Unc individuals lacking fluorescence. Worms were scored in late L3 after synchronization via egg prep. AC displacement was scored using Nomarski optics.

Effect of *RNAi* directed against genes required for fDTC migration: Worms were plated after L1 arrest, grown at 25°C, and scored 24-28 hours after feeding on plates seeded with feeding bacteria. All RNAi clones used came from a commercial library (Kamath et al. 2003), and RNAi targeting *lacZ* was used as a negative control. RNAi cultures were grown in 2xYT or LB media with 50µg/mL carbenicillin at 37°C overnight, and 70µL of each culture was added to room-temperature NGM plates with added 6uM IPTG and 100uM carbenicillin. Plates were allowed to dry for a minimum of two hours or up to overnight at room temperature before worms were added.

As the RNAi conditions affect fDTC outgrowth and hence gonad morphology, worms were staged based on age after plating, and scored for fDTC outgrowth defects at 63x using widefield microscopy, providing an internal positive control that RNAi had been effective. *lag-2p::rfp*

expression was used to identify the AC and P6.pxx nuclei. Photomicrographs were taken at 500ms exposure (*lag-2p::rfp*) and 750-900ms (*hlh-2prox::hlh-12::sl2::BFP*).

Acquiring potential bHLH transcripts for phylogenetic analysis

cDNA and/or protein sequences were obtained from the following sources:

For *C. elegans, C. briggsae,* and *C. remanei*: Wormbase versions WS266, WS277, and WS279 (wormbase.org)

For all other *Caenorhabditis* species except for *C. japonica* and *C. bovis,* as well as *D. coronatus, M. belari, O. tipulae,* and *P. oxycerus*: the Caenorhabditis Genomes Project v1 (caenorhabditis.org)

For *C. japonica*: Wormbase and the Caenorhabditis Genomes Project

For *C. bovis*: The Caenorhabditis Genomes Project and the European Nucleotide Archive (ebi.ac.uk/ena)

For *A. suum, H. bacteriophorae,* and *P. redivivus*: Wormbase ParaSite v9.0 (parasite.wormbase.org)

For *P. pacificus:* Wormbase, the Caenorhabditis Genomes Project, and Pristionchus.org

When available, entire genome sequences were downloaded for BLAST search using BLAST+ v.2.10.0 (ncbi.nlm.nih.gov). If genomes were not available for download, BLAST searches were performed online at the Caenorhabditis Genomes Project. Annotated *C. elegans* transcripts were BLASTed against both the cDNA and protein databases of other genomes to identify potential orthologs, using tBLASTx or tBLASTn to search all reading frames for translated sequences. All searches were done with an E value threshold of 10. WormBase version WS266, WormBase ParaSite v9.0, and the Caenorhabditis Genomes Project v1 were used to obtain

cDNA transcripts of BLAST results. All potential bHLH orthologs were screened using NCBI's Conserved Domain search (ncbi.nlm.nih.gov) against database CDD v3.17 – 55426 PSSMs with an E value threshold of 0.01, and transcripts without predicted bHLH domains were eliminated from the dataset (see Table S4.1).

Building and revising phylogenetic trees

Trees were generated using Bayesian analysis. cDNA sequences were grouped into FAS datasets using Sublime Text 3 (sublimetext.com). FAS datasets were translated and alignments built using MEGA7 (megasoftware.net) and ClustalW (amino acid) and exported as NEXUS files. NEXUS files were converted into XML format using BEAUti v1.8.4 (beast.community/beauty). Bayesian analysis was run on XML files either using BEAST 1.8.4 (beast.community) or BEAST2 with the BEAGLE library (beast2.org, github.com/beagle-dev/beagle-lib), or BEAST on XSEDE via the CIPRES Science Gateway (phylo.org), with a minimum of 3,000,000 generations.

BEAST output files were analyzed for the best-fit tree using TreeAnnotator 2.5.2, and resultant trees were evaluated using Tracer v1.7.1 (Rambaut et al., 2018). Trees with an effective sample size of under 200 according to Tracer were discarded. To generate the final trees, we used Tracer to determine the appropriate burn-in percentage for each tree, with a default of 10% burn-in, maximum clade credibility, and common ancestor for the node heights.

We used the full protein sequence for all of our trees, unless otherwise described, with HIF-1 orthologs serving as an outgroup for most analyses. As the initial BLAST search had very loose parameters, our initial tree drafts had several bHLH-domain-containing proteins that did not cluster with any known orthologs, which we hypothesized to be non-Class I or non-Class II genes. We

thus deleted these transcripts from our datasets and re-ran the trees to generate the final phylogenies.

BLAST searches for HLH-12 flanking regions

The bHLH domain of HLH-12 proteins from *C. elegans*, *C. japonica*, *C. briggsae,* and *C. castelli* was determined by both NCBI's Conserved Domain Search and UniProt KB (uniprot.org) when available. The flanking regions were used to perform both tBLASTx and tBLASTn searches as described above in *C. elegans, C. briggsae, C. japonica, C. castelli, C. bovis, C. parvicauda,* and *C. monodelphis*. Only non-*hlh-12* results were taken from each search.

Synteny

WormBase version WS266 was used to identify the 5 most proximal 5' and 3' genes to *C. elegans hlh-12*. BLAST searches against all species but *C. japonica* and *C. elegans* were done using BLAST+ or online searches as described above, with an E value threshold of 1. Reverse BLAST searches against *C. elegans* were done using WormBase version WS266 against the PRJNA13758 database, with an E value threshold of 1. WormBase version WS266 was also used to search for *C. japonica* orthologs. WormBase version WS279 and GBrowse were used to determine gene location in *C. elegans*, *C. japonica*, and *P. pacificus.* The Caenorhabditis Genomes Project v1 was used to analyse the genomic loci of *C. uteleia, C. castelli, C. quiockensis, C. virilis,* and *C. plicata.*

Table 4.1. Key resources table.

Reagent	**Source**
GS8958: *arIs51; arIs222 him-5(e1490)*	This paper
GS7849: *qIs90; arIs222 him-5(e1490)*	This paper
GS9150: *arSi45; arIs51; arIs222 him-5(e1490)*	This paper
GS9219: *arIs45; qI90; arIs222 him-5(e1490)*	This paper
GS9116: *arSi45; arIs222 him-5(e1490)*	This paper
GS9165: *ina-1(ar639); arIs222 him-5(e1490)*	This paper
GS9497: *arSi45; ina-1(ar639); arIs222 him-5(e1490)*	This paper
GS9306: *qyIs176; arIs131*	This paper
GS9308: *arSi45; qyIs167; arIs131*	This paper
him-8(e1489); fos-1(bmd138)	This paper
arSi45; him-8(e1489); fos-1(bmd138)	This paper
GS9500: *fos-1(ar105)/ tmC16[unc-60(tmIs1210)]*	This paper
GS9501: *arSi45; fos-1(ar105)/ tmC16[unc-60(tmIs1210)]*	This paper
GS8195: *arIs222 him-5(e1490)*	This paper
GS9084: *arSi45*	This paper

Results

See Table S4.2 for predicted bHLH code gene orthologs

Addition of HLH-12 to the AC compromises its functions but does not alter its fate

To perform the desired pro-AC→hDTC and pro-AC→LC transformations shown in Chapter 3, I generated an ectopic expression line for HLH-12 ("+HLH-12", see Figure 3.1 for structure and Figure 4.4A for expression pattern). To determine if addition of HLH-12 alone was sufficient to cause AC fate or functional defects, I assessed these animals for both expression of the AC marker *cdh-3p::gfp* and abnormal or absent vulvas which would suggest that AC functions have been compromised. The +HLH-12 cells expressed *cdh-3p::gfp* at 100% penetrance and did not express hDTC fate marker *ceh-22bp::venus*, suggesting that AC fate was retained (Figure 4.4B). However, the animals also had very compromised vulva formation: 100% had egg-laying defects, with ~90% having abnormal vulval eversion and the remaining 10% completely lacking a vulva (Figure S4.1A). This suggests that, while the AC retains its fate, its functions in uterine and vulval patterning have been compromised.

After inducing vulval precursor cell fate, a wild-type AC will invade through the basement membrane surrounding the gonad to physically contact the developing vulva (see Figure 1.3). This process requires the transcription factor *fos-1*, a direct target of HLH-2 homodimers which regulates genes necessary for the formation of invadopodia on the apical membrane of the AC (Schindler & Sherwood, 2011; Sherwood et al., 2005). Among others, *fos-1* regulates expression of the zinc metalloprotease *zmp-1*, which helps to dissolve the basement membrane that separates the somatic gonad from the vulval cells (Sherwood et al., 2005). I used the transgene *zmp-1p::mcherry-moesinABD* to visualize invadopodia, as it contains the moesin actin-binding domain and thus localizes to the actin-rich invadopodia (Schindler & Sherwood, 2011). The +HLH-12

ACs showed several defects in invasion. First, the number of invadopodia coming from the lateral membrane was significantly reduced, and there were also invadopodia projecting from the basolateral membrane, a hallmark of *fos-1(0)* ACs (Figure 4.4C; David Matus, personal communication). In addition, expression of an endogenously-tagged FOS-1::GFP allele was dim or absent in the +HLH-12 ACs (Figure 4.4D). Taken together, these results indicate that the addition of HLH-12 to the ACs results in ACs with greatly impaired invasion due to the downregulation of *fos-1*.

Ectopic expression of HLH-12 leads to hDTC/LC migration factor-dependent displacement of the AC from its normal position

While examining the +HLH-12 ACs, I noticed that the AC was often displaced from its usual position. To quantify this displacement, I measured its distance from the center of the descendants of the primary-fated vulval precursor cell, P6.p. P6.p divides twice to give rise to four granddaughters, collectively termed P6.pxx; the WT AC usually sits above the middle of these four cells during L3 (see photomicrograph in Figure 4.5A). For details on quantification and statistical analysis, see Figure 4.5A, Figure S4.2B and Material and Methods. Using *lag-2p::tagrfp* to mark both the P6.pxx and AC nuclei, I found that the +HLH-12 ACs were significantly displaced from their normal position (Figure 4.5A). In addition, they were displaced anteriorly or posteriorly at equal frequency. I also found that this displacement was purely horizontal, as there was no significant difference in vertical distance from P6.pxx (Figure S4.1B).

While these +HLH-12 ACs might be displaced due to some bHLH-independent mechanism, the most likely hypothesis is that the +HLH-12 ACs are actually migrating due to HLH-12 activating the machinery used for the hDTC/LC leader function. In the hDTC and LC,

HLH-12 directly activates both *gon-1*, a metalloprotease that dissolves the extracellular matrix to allow passage, and *ina-1*, an α-integrin that helps the cells adhere to the basement membrane and crawl forward (Blelloch & Kimble, 1999; Meighan et al., 2015). If the displacement of the +HLH-12 ACs does result from HLH-12-induced mobility, I would expect the +HLH-12 cells to both express these target genes and to cease migration in their absence.

To test this hypothesis, I looked at *ina-1*. I first generated an endogenously-tagged line using CRISPR, *ina-1(ar639)*, and examined its expression in the proximal gonad. Consistent with previous reporter transgenes (Meighan, Kann, et al., 2015; Meighan & Schwarzbauer, 2007), *ina-1(ar639)* is expressed in the hDTCs, LC, and the basement membrane surrounding the entire gonad; in L2, it is not expressed in the proximal gonad (Figure 4.5B, top panel). By contrast, I found expression of INA-1::GFP in the AC in L2 in roughly 50% of animals, suggesting that *ina-1* was ectopically expressed at this time (Figure 4.5B, bottom panel). In addition, I found that the +HLH-12 ACs were no longer displaced when I performed RNAi against *ina-1* (Figure 4.5C, Figure S4.2C-D), suggesting that their displacement is dependent on *ina-1*. Crucially, the WT ACs were not displaced by *ina-1* RNAi, indicating that the loss of *ina-1* does not result in AC movement due to lack of adhesion to the basement membrane. I also performed RNAi against *gon-1*, the other known direct target of HLH-12:HLH-2 dimers, but found an overall gonad bursting phenotype consistent with prior reports which impeded my displacement analysis (Figure S4.2A).

The combined observations that INA-1 is ectopically expressed in +HLH-12 ACs and displacement of +HLH-12 ACs is suppressed with RNAi against *ina-1* suggest that HLH-12 is ectopically activating INA-1 in these animals, leading to *ina-1*-dependent displacement of the AC in a manner consistent with *ina-1*'s role in promoting hDTC and LC migration. The hDTC and LC display characteristic features of migrating cell types such as focal adhesions and cytoskeletal

rearrangements (Wong & Schwarzbauer, 2012); *ina-1* is involved in focal adhesions, supporting the hypothesis that the addition of HLH-12 to the AC is sufficient to turn the AC into a migrating cell type. To further test this hypothesis, I examined two other genes which are not known to be direct HLH-2:HLH-12 heterodimer targets: *cdc-42*, a Rho GTPase which interacts with the *ina-1* cytoplasmic tail (Meighan, Kelly, Krahe, & Gaeta, 2015); and *pfd-1*, a chaperone protein which acts in tubulin biogenesis (Lundin, Srayko, Hyman, & Leroux, 2008). I found that RNAi directed against both *cdc-42* and *pfd-1* suppressed the displacement of the AC in +HLH-12 hermaphrodites (Figures 4.5C, S42B-D).

Finally, I examined AC displacement in *fos-1(0)* animals. The +HLH-12 ACs were still displaced, suggesting that the invasion defects described above do not contribute to AC displacement (Figure 4.5D). Overall, these results suggest that addition of HLH-12 to the ACs results in a cell that retains AC fate, but that expression of HLH-12 in these cells is sufficient to activate AC movement by the mechanism used for leader cell outgrowth. We propose also that invasion in these cells is impaired by the lower availability of HLH-2 homodimers due to the segregation of some HLH-2 in HLH-2:HLH-12 heterodimers, thus leading to the lowered expression of the HLH-2 homodimer target FOS-1 and impaired invasion as a result (Schindler and Sherwood, 2011; Figure 4.6).

hlh-2 and most bHLH code genes are conserved, but *hlh-12* is not

As *Ce-hlh-12* is sufficient to induce AC migration when expressed in the *C. elegans* pro-ACs, I next wanted to ask if the migrating AC in species like *M. belari* was due to ectopic *hlh-12* ortholog expression in the ACs of those animals (see Figure 4.3). bHLH proteins in general are very conserved in their sequence, binding motifs, and transcriptional targets (Massari & Murre,

2000). Among the genes making up the bHLH code, most have evidence for conservation as well. Maria Sallee found that two human orthologs of HLH-2, E47 and E12, were able to rescue *hlh-2(bx108)* defects and were unstable in VUs, suggesting strong functional and regulatory conservation (Sallee and Greenwald 2015). In addition, LIN-32 is a predicted ortholog of *Drosophila Atonal* (Portman and Emmons 2000), HLH-3 of *Achaete-Scute* (Doonan et al., 2008), and HLH-8 of *Twist* (Harfe et al., 1998); all highly conserved gene families (Valerie Ledent & Vervoort, 2001). Thus, I expected to find orthologs of the bHLH code genes in many nematode species outside *C. elegans*.

To find orthologs, I created a phylogenetic tree of HLH-2 and Class II bHLH genes across various nematode species (Figure 4.8A; see also Figure 4.7A and Figure S4.3). The tree included all annotated orthologs of *C. elegans* Class I and Class II bHLH genes from *C. japonica* and *P. pacificus,* as well as bHLH-domain containing BLAST hits from 7 other species chosen to give a good representative spread across nematode clades III-V (David Fitch and Karin Kiontke, personal communication; see also Figure 4.7A). I used the Class VII bHLH gene *hif-1* as an outgroup for this tree. As *M. belari* was of particular interest to us due to its migrating AC (Félix & Sternberg, 1996), we used the more distant species *P. oxycerus, P. redivivus* and *A. suum* to serve as outgroups in the individual gene clades.

For the clades formed by orthologs of the bHLH code genes, see Figure 4.8B. As expected, I found that most of the bHLH code genes do indeed show strong conservation. HLH-2 in particular is highly conserved, with a predicted ortholog in each of the 10 species contained in the tree. The HLH-2 clade contains all annotated *hlh-2* transcripts and is arranged similarly to the predicted evolutionary relationship between the species, with transcripts from the included *Caenorhabditis* species clustering together and transcripts from the more distantly-related species

such as *A. suum* and *P. redivivus* forming a different clade. The LIN-32 clade is similar, with representation from all 10 species in the tree and general conservation of their phylogenetic relationships. Of note, there are two representatives from *P. oxycerus*, suggesting a potential gene duplication in this lineage.

The HLH-8 clade is similar to those of HLH-2 and LIN-32, containing all annotated *hlh-8* transcripts and with transcripts that cluster in a predictable order. However, the HLH-8 clade contains transcripts from only 8 species, as no HLH-8 ortholog could be identified in *M. belari* and *H. bacteriophora.* As there are predicted orthologs in lineages both more basal and more evolved, this suggests either two lineage-specific losses of *hlh-8* or a lack of sufficient quality data in the draft genomes for these species.

The HLH-3 clade is more complex than those for HLH-2, HLH-8, and LIN-32, as it is mixed with predicted HLH-14 orthologs. However, previous studies (Grove et al., 2009) suggest that *hlh-3* and *hlh-14* are very closely related, lending further support to this mixed clade, and a recent re-release of the *P. pacificus* genome with updated annotations contains only a single annotated *Ppa-hlh-3/hlh-14*. Taken together, the HLH-3/HLH-14 clade contains all six annotated HLH-3 and HLH-14 transcripts and at least two transcripts from each species, suggesting that HLH-3 and HLH-14 are both conserved as well.

Strikingly, the clade for HLH-12 contains only the two annotated transcripts from *C. elegans* and *C. japonica* (*P. pacificus* had no annotated *hlh-12* transcript), and no representatives from any other species. This suggests that, unlike the other bHLH code genes, HLH-12 is not conserved.

hlh-12 likely arose within *Caenorhabditis*

Of the 10 species tested above, I found HLH-12 orthologs in *C. elegans* and *C. japonica*, but none in species outside *Caenorhabditis*. Interestingly, I also found no predicted ortholog in *C. monodelphis*, the most basal lineage in *Caenorhabditis* (Slos et al., 2017). This raises the possibility that *hlh-12* arose at some point within the *Caenorhabditis* lineage.

To determine which lineages within *Caenorhabditis* have *hlh-12* orthologs, I performed a phylogenetic analysis as described above on *Caenorhabditis* species. This analysis includes *Diploscapter coronatus,* a member of the *Caenorhabditis* sister genus*,* as well as *Oscheius tipulae* to serve as an outgroup for each ortholog clade. As a phylogeny of all *C. elegans* bHLH genes showed that *hlh-11* and *hlh-1* are in the same clade as *hlh-2* and all Class II bHLH genes (see below and Figure 4.10), I also included these in my analysis. In addition, previous work from Grove et al. (2009) suggested that the Class VII bHLH-PAS protein HIF-1 is the most closely related protein to CEL-HLH-12 (see Figure 4.2); I thus added predicted HIF-1 to this tree as well. I used a subset of *Caenorhabditis* species chosen for a good spread across the clade, including *C. bovis,* a recently-described basal species (David Fitch and Karin Kiontke, pers. comm.; Stevens et al., 2020; see Figure 4.7B).

For the complete tree of all *Caenorhabditis* species except for *C. bovis*, see Figure S4.4; for the tree of selected *Caenorhabdits* species including *C. bovis,* see Figure S4.5. I found that *C. bovis* is the most basal species with a predicted *hlh-12* ortholog, and that, with a few exceptions, *hlh-12* orthologs are present in all higher *Caenorhabditis* species (Figure 4.9). The absence of *hlh-12* in *Diploscapter* further supports the hypothesis that *hlh-12* originated with *C. bovis,* and thus is unique to a subset of *Caenorhabditis* species.

How *hlh-12* evolved remains unclear

The apparent origin of *hlh-12* in *Caenorhabditis* raises the question of how the gene evolved. One obvious hypothesis is via duplication of another bHLH gene. Another hypothesis is that *hlh-12* evolved through domain shuffling, which has been suggested as a probable mechanism for bHLH gene evolution (Morgenstern & Atchley, 1999).

To assess if *hlh-12* might have evolved through gene duplication, I performed a phylogenetic analysis of all *C. elegans* bHLH genes (Figure 4.10). Using only genes from *C. elegans* allowed me to more clearly visualize relationships between the genes, and I included all bHLH genes in case *hlh-12* had arisen from a duplicate of a non-Class I or Class II gene. In addition, I used only the bHLH domains in this analysis, to avoid confusion from any extra domains such as the PAS domains in Class VII bHLH proteins.

Of note, my analysis did not show a relationship between the HLH-12 and HIF-1 bHLH domains, contrary to previous findings (Grove et al., 2009). This phylogenetic analysis did show close relationships between *hlh-3/14* and *hlh-17/31/32*, suggesting that these genes arose via duplication; however, *hlh-12* has no obvious sister, instead appearing in a clade with *hlh-3, hlh-14, hlh-4,* and *hlh-1.* While this does not rule out gene duplication completely, it does suggest that gene duplication is unlikely in this instance—for example, my analysis suggests that *hlh-3* and *hlh-14* split before *H. bacteriophora* and *O. tipulae*, as both of these species have predicted separate *hlh-3* and *hlh-14* orthologs (Figure 4.8B); however, *hlh-12* is more different from its closest sister *hlh-4* in this tree than *hlh-3* and *hlh-14* are from each other, suggesting that if there was a gene duplication event it happened before the *hlh-3/hlh-14* split. As I predict that *hlh-12* arose within *Caenorhabditis*, at a point much higher up the lineage than the *hlh-3/hlh-14* split, gene duplication as an explanation for *hlh-12*'s origin seems unlikely.

Another hypothesis for how *hlh-12* evolved is through domain shuffling, the process by which an exon containing a functional domain is spliced into the intron of another gene. As Class II bHLH genes in particular contain only the bHLH functional domain, this provides an attractive explanation for how they might evolve (Morgenstern & Atchley, 1999). Supporting this hypothesis is the fact that *Cel-hlh-12* is contained within an intron of another gene, *spp-10* (see Figure 4.11).

To determine if *hlh-12* evolved through domain shuffling, I performed a BLAST search against the N- and C- terminal bHLH-domain flanking regions of HLH-12 orthologs from various *Caenorhabditis* species. If *hlh-12*'s bHLH domain did come from an exon spliced into an intron at another locus, these flanking regions should still retain some signature of their original location shared across these species. However, I found that the BLAST results did not suggest a common location (see Table 4.5). While my search was likely hindered by the fact that many of these sequences are very small, the complete lack of accord in the BLAST results suggests that *hlh-12* did not arise through domain shuffling. In addition, the *hlh-12* orthologs in non-*C. elegans* species are not contained within introns (see Figure 4.11), providing further evidence that *hlh-12* arose via some other method.

Taken together, my results suggest that *hlh-12* did not arise through two common forms of gene evolution, gene duplication and domain shuffling. I finally turned to a synteny analysis to determine if the genomic locus of *hlh-12* might provide some evidence for how it evolved. I took the 5 most proximal 5' and 3' genes in *C. elegans,* as well as *spp-10,* which contains *Cel-hlh-12* in an intron, for a total of 11 genes of interest, and looked at their conservation and location in other *Caenorhabditis* species (Figure 4.11). I found that these genes had varying levels of conservation, with *spp-10* in particular being highly conserved and others like *ant-1.4* appearing more novel.

The *hlh-12* loci in *C. briggsae* and *C. japonica* resembled the *C. elegans hlh-12* locus, containing many of the same genes with a roughly conserved order. In particular, *spp-10* was always close to *hlh-12.* However, I found that in *C. castelli,* one of the most basal species containing *hlh-12,* the *spp-10* and *hlh-12* loci were different, with the *spp-10* locus also containing two of the other proximal genes (Table 4.6). As the genomes for many *Caenorhabditis* species are arranged in relatively small scaffolds, I was unable to ascertain the position of the *spp-10* locus relative to the *hlh-12* locus, but I did ascertain that the *hlh-12* locus contained no orthologs of any of these proximal genes within 41 kilobases 5' and 9 kilobases 3' of *Cca-hlh-12* (Table 4.6). Other more basal species also had this split between *hlh-12* and *spp-10* loci, with the *spp-10* locus containing orthologs of the proximal genes. Taken together, this suggests that *hlh-12* evolved in a different place in the genome, and translocated to its current locus at some point more recently, most likely with *C. japonica* (Table 4.6, Figure 4.9, Figure 4.11).

Discussion

Addition of HLH-12 to the pro-ACs results in ACs with compromised later functions

ACs with added HLH-12 migrate using the hDTC/LC migration mechanisms, but retain AC fate. In addition, I found that AC invasion was impaired, with ~65% of animals showing a reduction in FOS-1 expression and a much lower average number of invadopodia overall. I hypothesize that this is due to a titration model, in which the added HLH-12 results in less HLH-2 available to form homodimers, and therefore compromise of HLH-2 homodimer-dependent functions such as invasion (see Figure 4.6).

As discussed in Chapter 1, HLH-2 regulates both AC fate specification and all of its functions, including induction. While I do see some failure of induction in the +HLH-12 AC

animals, it is only at a penetrance of around 10%, much lower than the penetrance of downregulated FOS-1 (Figure 4.4D, Figure S4.1A), and I see no failure in AC specification. Why, then, do I see such a compromise of invasion, but minor or no defects with specification and induction? One possible explanation is that HLH-12 requires some time to build up levels from the relatively weak *hlh-2prox* promoter, and that at the time of specification and induction there is still sufficient HLH-2 homodimer for those processes; another explanation is that *fos-1* is more sensitive to HLH-2 homodimer levels than the *lin-3*/EGF that serves as the induction signal. One way to test this would be to drive HLH-12 from a promoter that expresses later in the AC, such as *cdh-3p* or *zmp-1p*, and see if this later addition of HLH-12 still impairs invasion.

Expansion and lineage-specific losses of bHLH genes within nematodes

As expected, I found much conservation in the Class I and II bHLH genes used in our analysis. In particular, I found HLH-2 orthologs in each of the 10 species included in the original tree, and in all but *C. latens* in the *Caenorhabditis*-specific tree (see Table S4.2). The apparent lack of *hlh-2* in *C. latens* could be due to incomplete genome data, as other studies suggest that Class I/E proteins are highly conserved (Massari & Murre, 2000; Sallee & Greenwald, 2015; Zhuang et al., 1998). In general, however, I find clearly predicted HLH-2 orthologs in each of the species used in our analysis, further supporting its deep conservation. *P. pacificus* is known to have two *hlh-2* orthologs, *Ppa-hlh-2.1* and *Ppa-hlh-2.2*(Rödelsperger et al., 2017); in addition, I found evidence for gene duplication of *hlh-2* in *C. brenneri* and *C. uteleia* (Table S4.2, Figure S4.4), suggesting that other species might also possess more than one Class I bHLH gene.

A more striking finding was the apparent expansion of Class II factors throughout Nematoda. While *C. elegans* has 18 predicted Class II bHLH genes (Grove et al., 2009), I found

predicted orthologs for much fewer Class II transcripts in all non-annotated, non-*Caenorhabditis* species included in our tree (Table S4.2). Our results suggest that the least conserved genes are *hlh-31* and *hlh-32*, which we could not find outside *C. elegans*, and *hlh-12, hlh-17* and *hlh-19*, which we could not find outside *Caenorhabditis* (Table S4.2).

Of the species studied in this analysis, *P. pacificus* is particularly interesting because these preliminary findings suggest that there may be a number of species- or genus-specific losses of Clas II bHLH genes. Interestingly, a recent re-annotation of the *P. pacificus* genome shrunk the number of predicted Class II orthologs, from the 11 initially predicted to eight, including a single *Ppa-hlh-3/hlh-14* in line with its predicted duplication in later lineages (Rödelsperger et al., 2017). My analysis found evidence of potential *Pristionchus* losses of *hlh-10, hlh-13,* and *hlh-15*, as I found predicted orthologs for all of these genes in lineages both below and above *P. pacificus* (Figure S3). These apparent losses are somewhat surprising, as both *hlh-13/*PTF1 and *hlh-15/*NSCL are the only predicted nematode orthologs of conserved bHLH gene families (Table 4.3; Liachko et al.; 2009; Mansfeld et al., 2015). Both *hlh-13* and *hlh-15* have also implicated in aging and the dauer process: *hlh-13* as a direct target of *daf-16,* and *hlh-15* as a regulator of branched-chain amino acids during the aging process (Liachko, Davidowitz, & Lee, 2009; Mansfeld et al., 2015). If these losses are true, this would be an interesting finding, as it would suggest that *P. pacificus* or *Pristionchus* species in general are missing many bHLH genes that are otherwise quite conserved, especially those regulating dauer and the aging process. A more careful phylogenetic analysis focused on species closely related to *Pristionchus* would be necessary to investigate this further.

Several bHLH genes, appear to be novel to *Caenorhabditis* or subsets of the clade

I do not find *hlh-12*, *hlh-17, hlh-19, hlh-31,* and *hlh-32* outside *Caenorhabditis* (Table S4.2, Figure S4.4). The latter four genes were all singled out by Grove et al. (2009) as those grouped in Class II because of sequence similarities but not dimerization data. However, we find using the orthology prediction tool Ortholist 2 that the predicted human orthologs of *hlh-17, hlh-31,* and *hlh-32* are the E proteins BHLE22 and BHLE23, instead of other Class II factors (Kim, Underwood, Greenwald, & Shaye, 2018). This disparity could result from Ortholist assigning orthology based on the entire sequence, while Grove et al. used only the bHLH domain in their work; however, we did not find that any annotated *hlh-17, hlh-19, hlh-31,* or *hlh-32* transcripts clustered with *hlh-2*, the known E protein ortholog used in our analysis, and thus we do not believe them to be true Class I proteins.

Of the four, Grove et al. found that *hlh-17, hlh-31,* and *hlh-32,* all Olig orthologs, formed a single clade; our results support that finding, and we predict that *hlh-31* and *hlh-32*, which we find in *C. elegans* only, resulted from duplications of *hlh-17* (Table S4.2, Figure S4.4, Figure S4.5). *hlh-19* is an interesting case, as various sources disagree on its classification: Vervoort and Ledent classify it as an Achaete-Scute, while Grove et al. classify it as an ortholog of Hand, and Ortholist provides no known human orthologs. Given its unclear status, it is not surprising that *hlh-19* does not appear to be highly conserved, appearing instead to be a novel feature of the Elegans supergroup of *Caenorhabditis* based on our analysis (Figure 4.7B, Figure S4.4).

hlh-12's origins remain a mystery

In my analyses above, I found that *hlh-12* evolved with *C. bovis*, and provided evidence suggesting that it did not evolve through either gene duplication or domain shuffling (see Figure

4.10, Table 4.6). Gene duplication and domain shuffling are only two of many mechanisms by which genes can arise: others include gene fusion, gene fission, and other combinations of duplication and recombination (Long et al., 2013). Any one or several of these mechanisms could have contributed to *hlh-12*'s origin.

If *hlh-12* did arise from the genetic material of some other gene, such as through gene fission or fusion, it remains likely that another bHLH gene was the source given *hlh-12's* close phylogenetic relationship with other Class II bHLH genes in particular (Figure 4.10). While many of the Class II bHLH genes have clear orthologs in other species, such as *hlh-8/Twist* or *hlh-3/Achaete-scute*, previous work has unsurprisingly disagreed on the orthology of *hlh-12*. I find that different analyses predict different relatives: for example, my analysis of *C. elegans* bHLH domains showed that *hlh-12* was most closely related to the Achaete-scutes *hlh-3/4/14*, while my analysis of all Class I and Class II bHLH genes in *Caenorhabditis* and *Diploscapter* using the whole protein sequence found that *hlh-12* was most closely related to *hlh-10* (Figure S4.3, S4.4). These analyses differ in both included species and sequences, with a major difference being the inclusion or absence of non-bHLH-domain sequence. Thus, focusing future analyses on the bHLH domain only, such as an analysis of *Caenorhabditis* and *Diploscapter* species in this manner, might provide further evidence about whether or not other bHLH genes could have contributed to *hlh-12*'s origin.

During my synteny analysis, I discovered that *hlh-12* appears to have originated at a different locus than its current position in *C. japonica* and above species, including *C. elegans* (Figure 4.11, Table 4.6). The locus of *spp-10*, the gene containing *Cel-hlh-12*, is much more conserved, further suggesting that *hlh-12* originated elsewhere and only recently moved to its current location in the higher *Caenorhabditis* species. Further syntenic analysis of the separate

hlh-12 locus in species such as *C. castelli* and *C. bovis* might shed some light onto how *hlh-12* evolved there. In addition, if further phylogenetic analyses suggest a stronger candidate for a bHLH gene parent of *hlh-12*, investigation of its syntenic locus might also provide useful information about how *hlh-12* originated.

Potential for a bHLH code in other nematode gonads?

With the striking exception of *hlh-12*, the bHLH code genes were highly conserved across all species studied. This high level of conservation, and in particular the absolute conservation of *hlh-2*, supports the idea that the bHLH code could be conserved across species; in addition, the fact that addition of CE-HLH-12 to the *C. elegans* AC resulted in a change of function provides a mechanism by which other species could have evolved different gonad morphologies. As to date there are no known conserved mechanisms of gonad development in nematodes, and indeed predicted gene networks for essential processes such as induction and vulva formation vary across species (Félix & Sternberg, 1997; Kiontke et al., 1925), the conservation of the bHLH code across species would be a truly novel finding.

While the lack of *hlh-12* orthologs outside *Caenorhabditis* suggests that *hlh-12* orthologs cannot be controlling the migration of regulatory cells in all species, it is possible that other Class II bHLH proteins are dimerizing with HLH-2 to target the same effector genes. Supporting this hypothesis, I have found preliminarily that *Ce-hlh-12* target genes *ina-1* and *gon-1* are highly conserved, and that *C. elegans gon-1* and *ina-1* have E boxes in conserved areas of their 5' intergenic regions. *Ce-hlh-12* shares an E box consensus motif with *Ce-hlh-3, Ce-hlh-14, Ce-hlh-19,* and *Ce-lin-32* (Grove et al., 2009; Narasimhan et al., 2015); as E boxes are generally very conserved, with both HLH-2 and E2A dimers recognizing the same sequence, this would mean

that potentially orthologs of any of the listed genes could substitute for the absent *hlh-12* ortholog activity (Grove et al., 2009; Massari & Murre, 2000).

The general conservation of bHLH genes in other species support the hypothesis that bHLH genes are regulating their gonads, but at the moment I do not know if bHLH genes are expressed in the gonad of any species outside *C. elegans*. The first step to testing this hypothesis would be to look for HLH-2 expression in the gonad of a species like *P. pacificus*, which has the advantage of a high-quality genome sequence and clearly identified regulatory cells, using tools like CRISPR or smFISH (Barkoulas et al., 2016; Witte et al., 2014). If HLH-2 is present in the regulatory cells, that would support the hypothesis that the code is active in other species; from there, looking at expression of the Class II bHLH genes in *P. pacificus* or perhaps testing their function in a model like the *C. elegans* AC could suggest what other bHLH genes, if any, might be contributing to regulatory cell specification and function in other species.

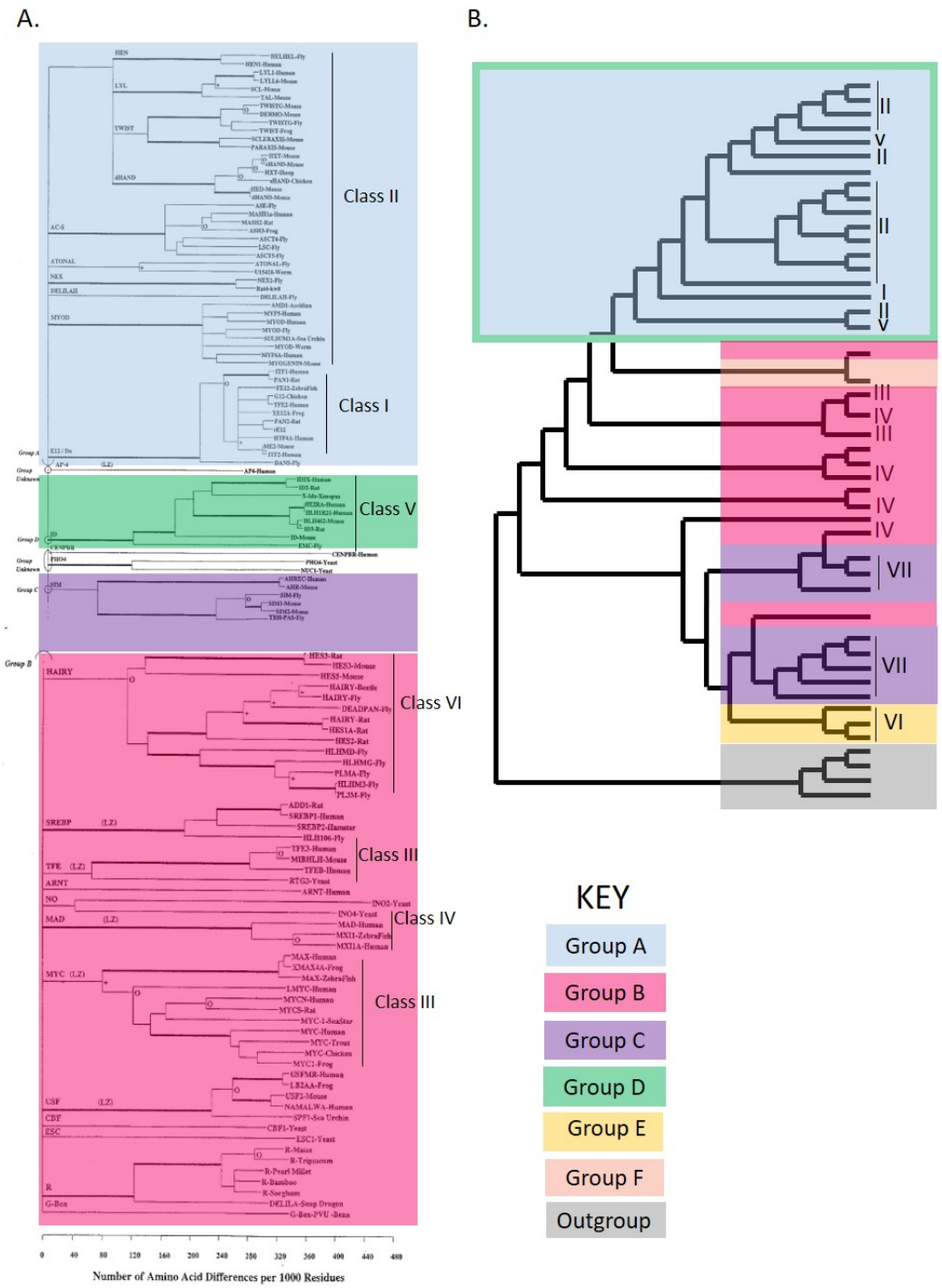


Figure 4.1. Comparison of previous bHLH phylogeny studies. Colors indicate groups as first suggested by Atchley and Fitch; classes according to the system from Massari and Murre are indicated next to lineages. In B, blue box with green outline represents a clade with mixed group A and group D members as according to Atchley and Fitch (1997). Adapted from Atchley & Fitch, 1997; Ledent & Vervoort, 2001; Massari & Murre, 2000.

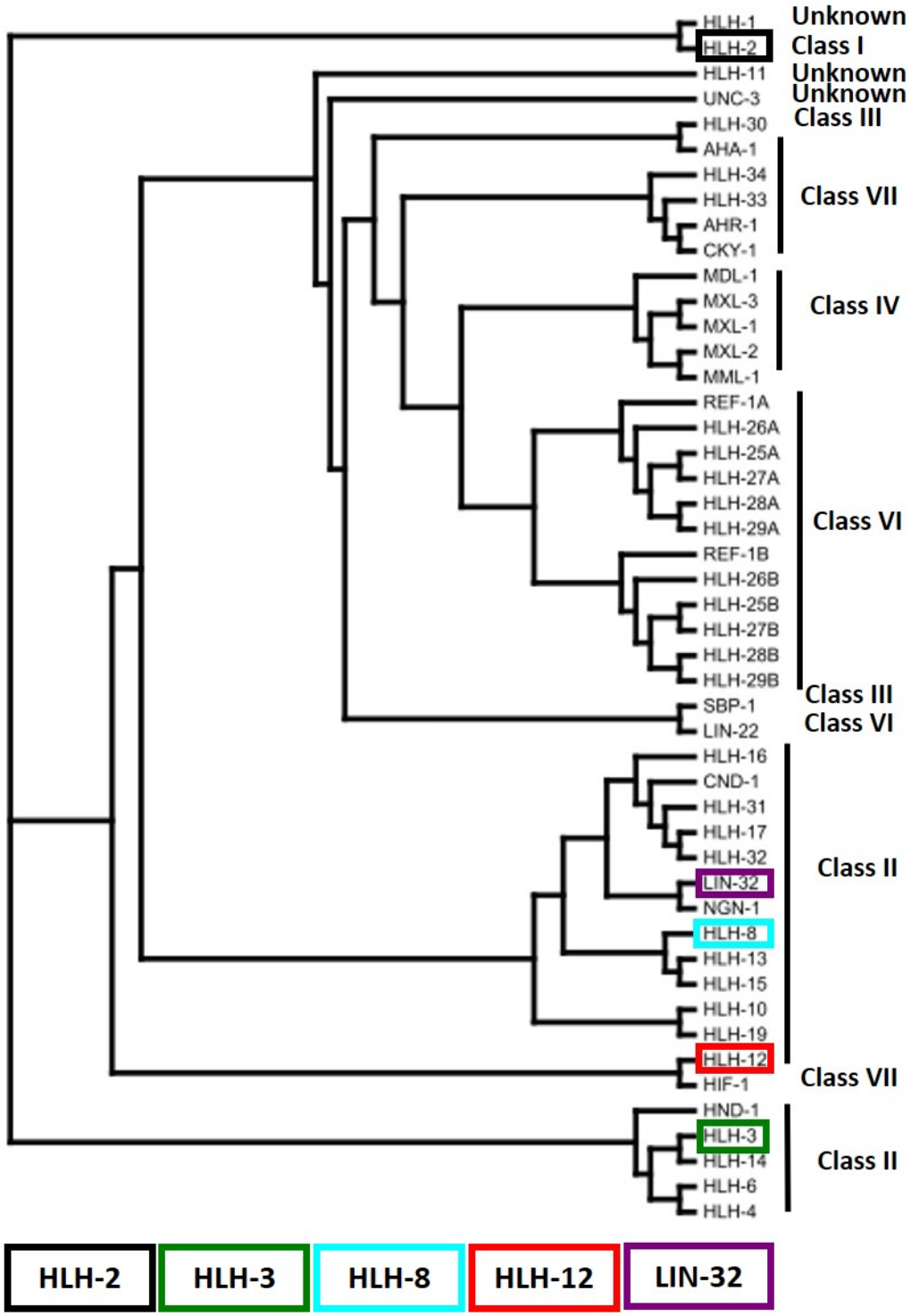


Figure 4.2. Phylogeny of *C. elegans* bHLH proteins. Classes according to Massari and Murre are indicated to the right of each lineages. Colored boxes indicate bHLH code proteins. Adapted from Grove et al., 2009.

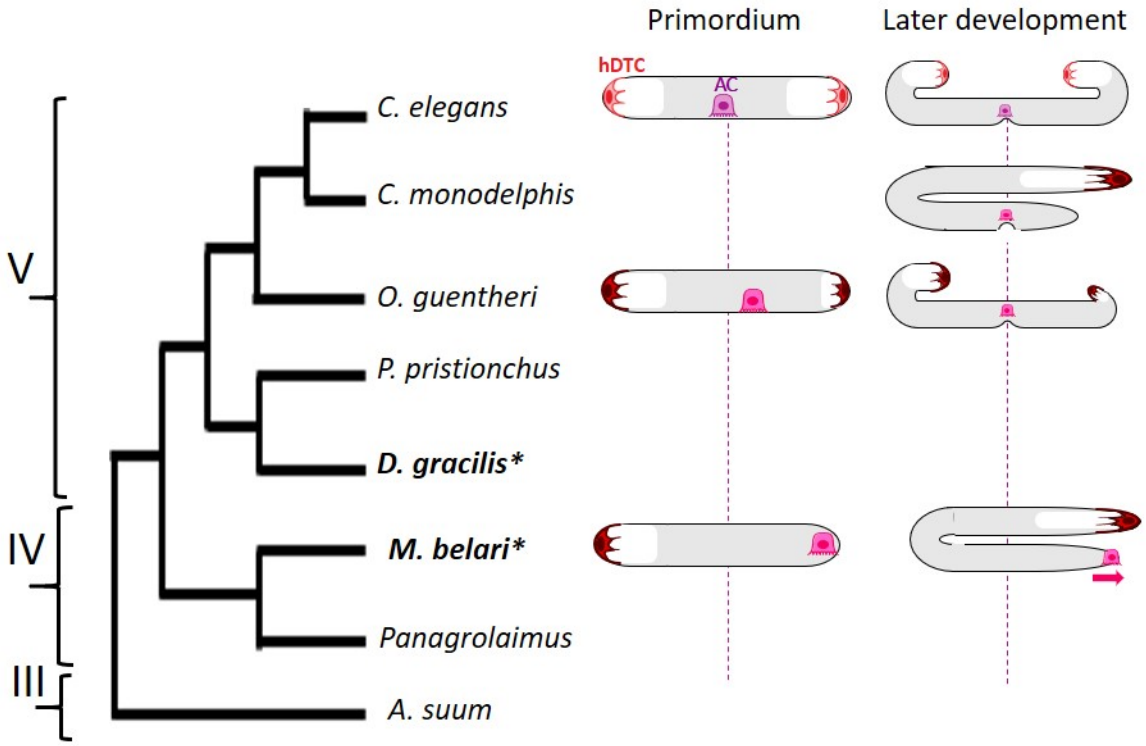


Figure 4.3. Introduction to nematode phylogeny and gonad development. Phylogeny not to scale. Nematode clades are indicated to the left. Species with bold and * indicate gonad-independent induction of VPC fate. Gonad development shown at primordium and L4 stages; white boxes indicate forming germline, and pink arrow in *M. belari* indicates direction of AC movement. Dashed lines indicate position of central AC, as in *C. elegans*. Adapted from Félix & Sternberg, 1997; Haag, Fitch, & Delattre, 2018; Haag, Helder, Mooijman, Yin, & Hu, 2018; Kiontke et al., 2007; Schulze & Schierenberg, 2011.

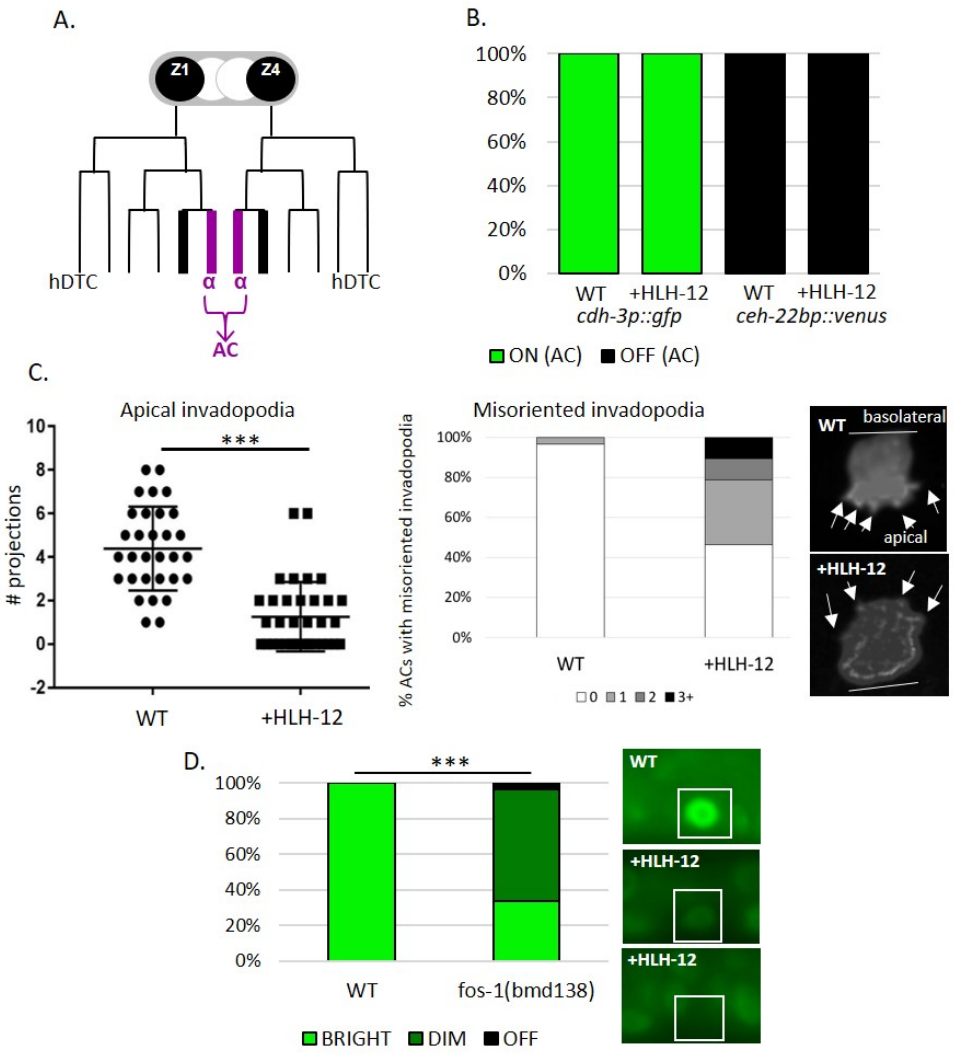


Figure 4.4. Ectopic expression of HLH-12 in the AC does not change its fate but compromises its function.

A. Expression pattern of *hlh-2prox*. Lineages expressing *hlh-2prox* are shown in bold; the lineages that give rise to the AC are shown in purple.

B. +HLH-12 ACs express the AC marker *cdh-3p::gfp*, and not the hDTC marker *ceh-22bp::venus*. n = 20-28 for each genotype.

C. +HLH-12 ACs have fewer invadopodia from the apical membrane and more from the basolateral membrane. n = 28-35 for each genotype. Fluorescent markers shown in photomicrographs is *zmp-1p::mcherry::moesinABD*; images are orthogonal projections. Arrows indicate invadopidia; line indicates lack of invadopodia on the indicated membrane. *** $p<0.001$, Kolmogorov-Smirnov. D. GFP::FOS-1 expression is downregulated in +HLH-12 ACs. White boxes surround AC. n = 20-30 for each genotype. *** $p<0.001$, Fisher's Exact T Test.

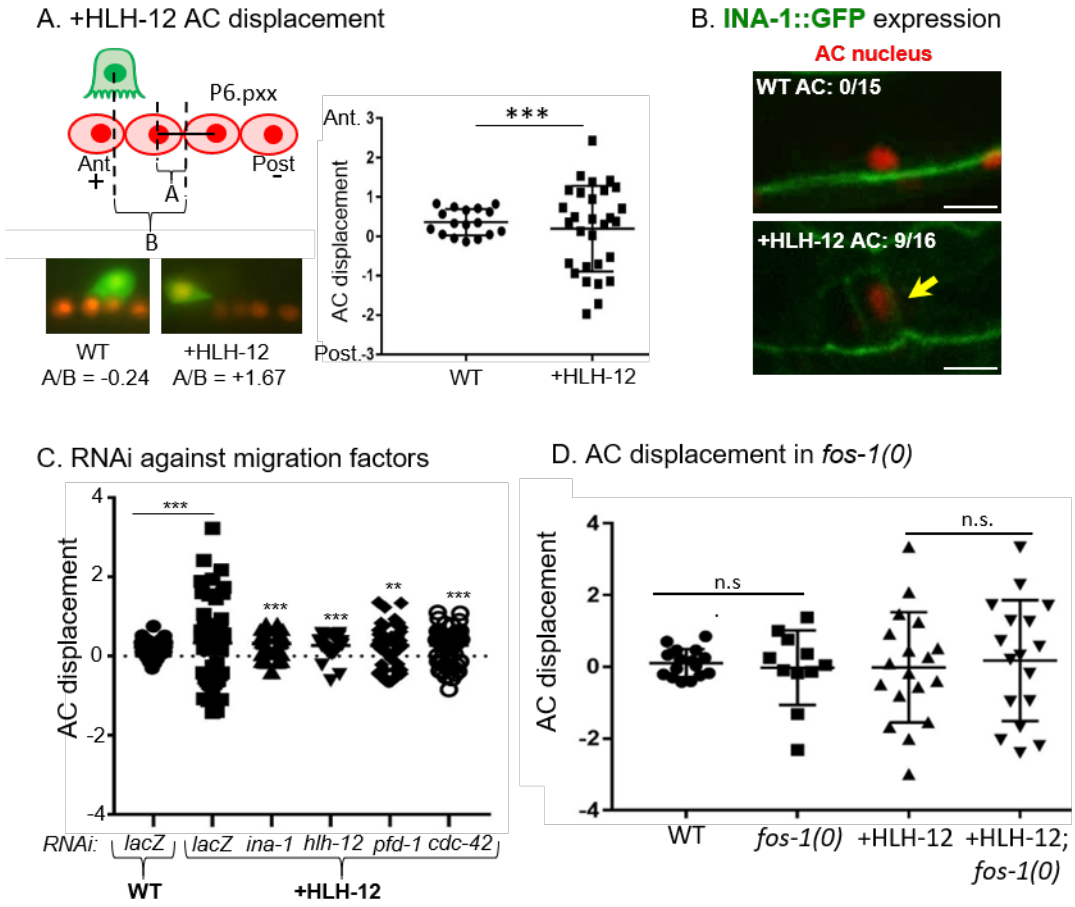


Figure 4.5. +HLH-12 ACs are displaced in an hDTC/LC migration factor-dependent manner.

A. Quantification of AC displacement; for more details, see Materials and Methods. AC in representative photomicrographs is marked with *cdh-3p::gfp,* and P6.pxx with *lag-2p::tagrfp*. n = 18 for WT and 27 for +HLH-12. ***p<0.001, Kolmogorov-Smirnov.

B. INA-1::GFP is expressed in the basement membrane, and on the cell membrane of the WT hDTC and LC (not shown). In L2, it is not expressed in the proximal gonad(top), but is expressed ectopically in the +HLH-12 AC (bottom, yellow arrow). n refers to animals with INA-1::GFP expression in the proximal gonad.

C. Knocking down factors required for hDTC and LC migration prevents AC displacement. n = 20-52 for each genotype andcondition. **p<0.01, ***p<0.001; Kolmogorov Smirnov. For information on quantification and statistics, see Materials and Methods and Figure S2B; for controls for displayed experiment and second RNAi trial, see Figure S2C-D.

D. +HLH-12 ACs remain displaced in the *fos-1(0)* background. n = 18-22 for each condition. n.s. Komogorov-Smirnov.

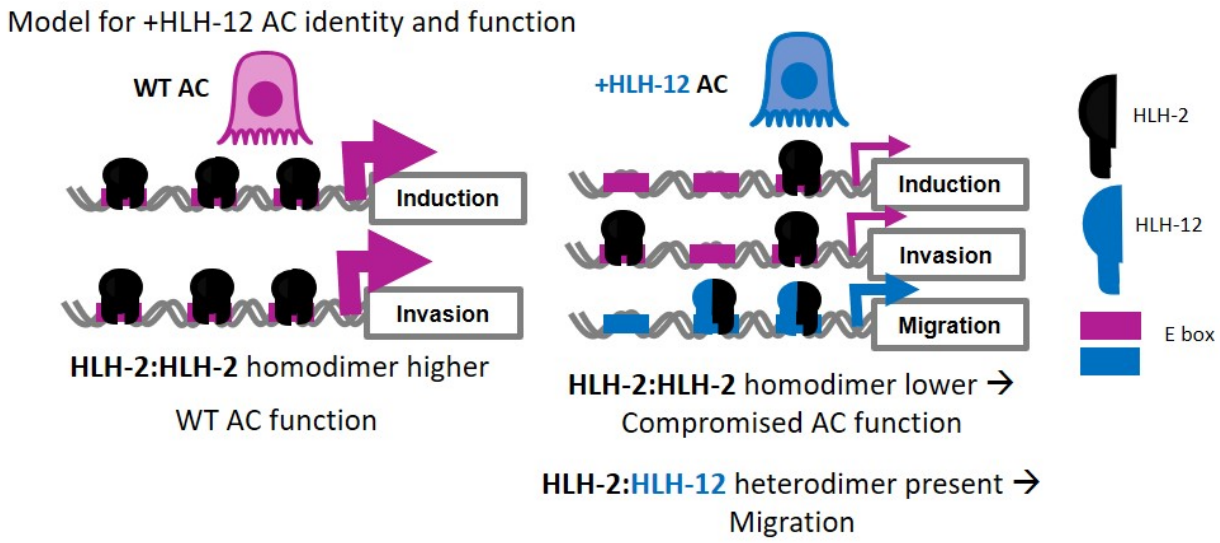


Figure 4.6. A model for +HLH-12 AC identity and function. In the WT AC (left), HLH-2 homodimers promote AC fate and functions like induction and invasion. In the +HLH-12 AC, addition of HLH-12 results in less HLH-2 available for homodimers and causes impairment in induction and invasion; meanwhile, HLH-12 heterodimers promote migration. Arrow size implies relative promotion of each function.

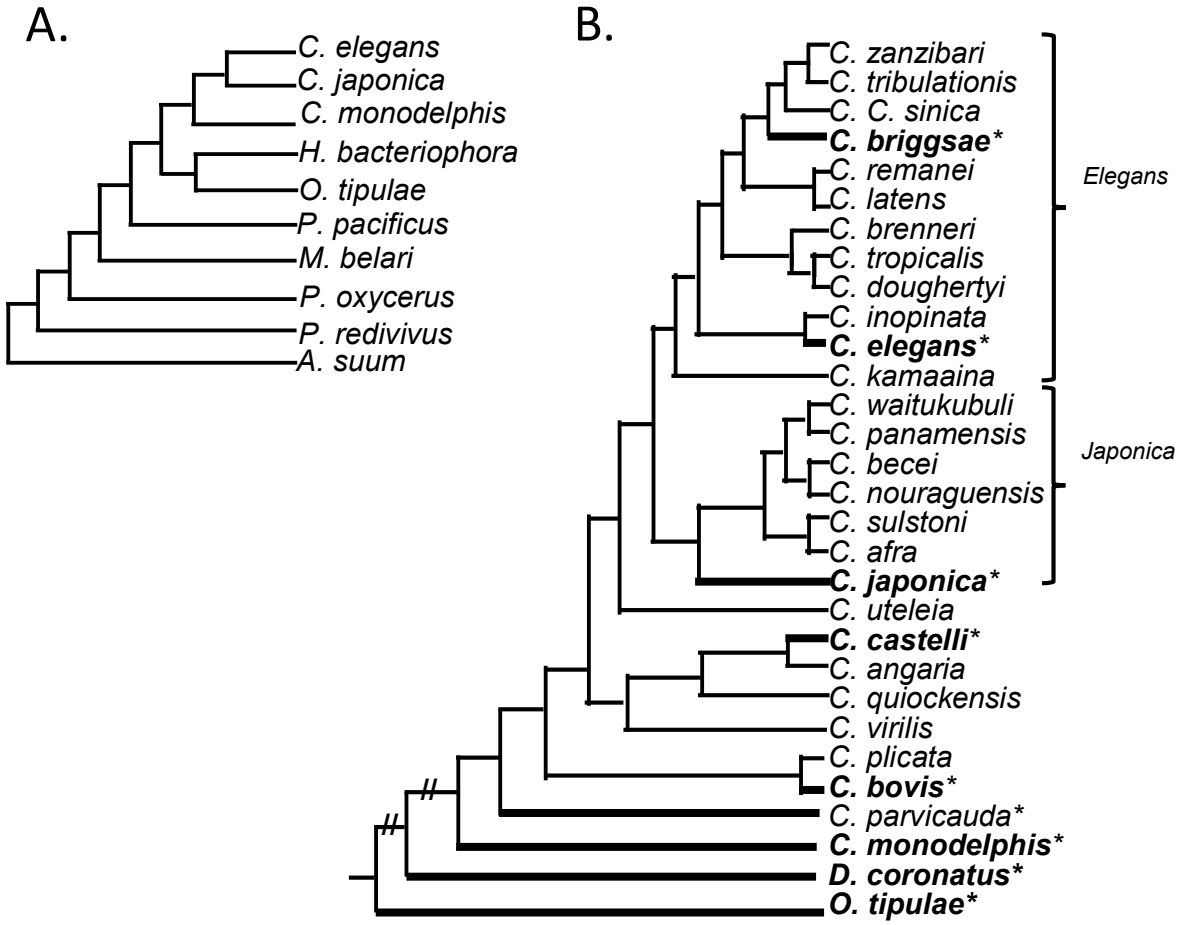


Figure 4.7. Species used for phylogenetic analysis. Phylogenies of species included in overall (A) and *Caenorhabditis*-only (B) analyses. Evolutionary distances are not to scale. Dashed lines in B indicates that species are not sisters of the above taxa. In B, species in bold and with * are those included in the *Caenorhabditis* + *Diploscapter* analysis; supergroup designations are shown to the right. A: adapted from Karin Kiontke and David Fitch, pers. comm.; B: adapted from Stevens et al., 2020.

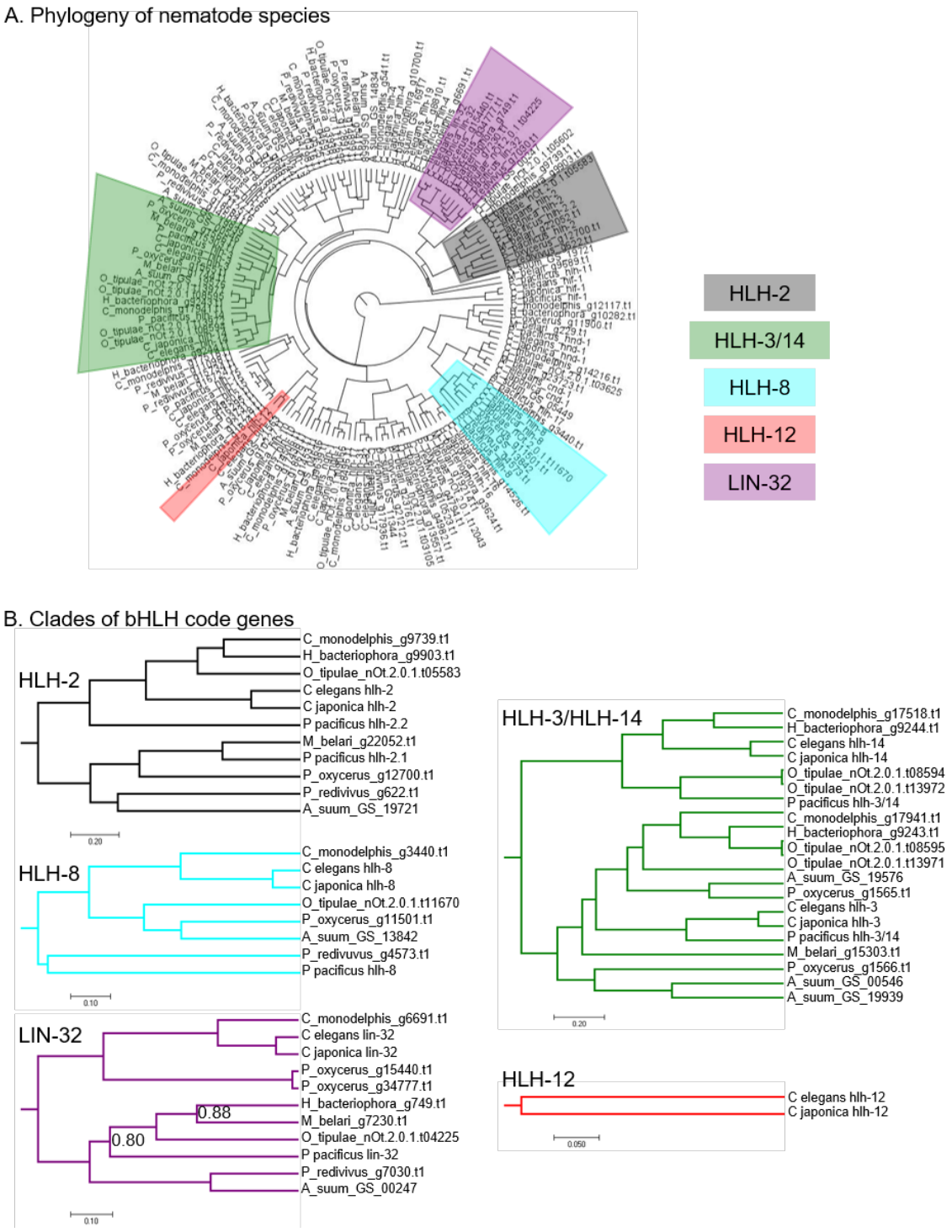


Figure 4.8. bHLH code clades predicted using 10 nematode species. See Figure S4.3 for larger full tree.

A. Colored segments indicate clades containing bHLH code gene orthologs.

B. Only posterior probabilities <0.9 are shown**.** Branch lengths are in substitutions/site.

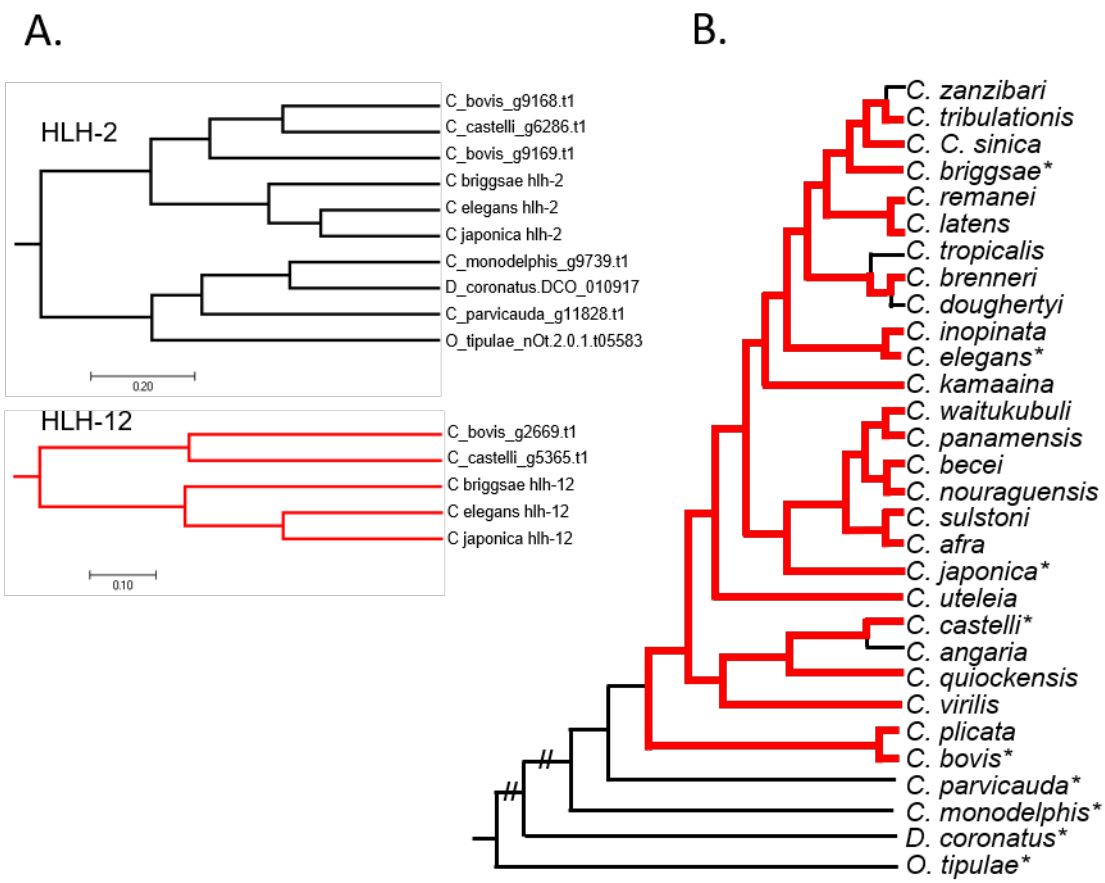


Figure 4.9. Predicted orthologs of HLH-12 in all *Caenorhabditis* species plus *Diploscapter coronatus*. For full phylogenies, see Figures S4.4 and S4.5.

A. bHLH ortholog clades for HLH-2 (top) and HLH-12 (bottom). All posterior probabilities are >0.9; branch lengths are in substitutions/site.

B. Phylogeny of *Caenorhabditis* and *Diploscapter* species, with those containing predicted CE-HLH-12 orthologs marked in red. Dashed lines indicate that species are not sisters of the above taxa. Phylogeny is not to scale. * indicates species analysed in smaller *Caenorhabditis* + *Diploscapter* tree.

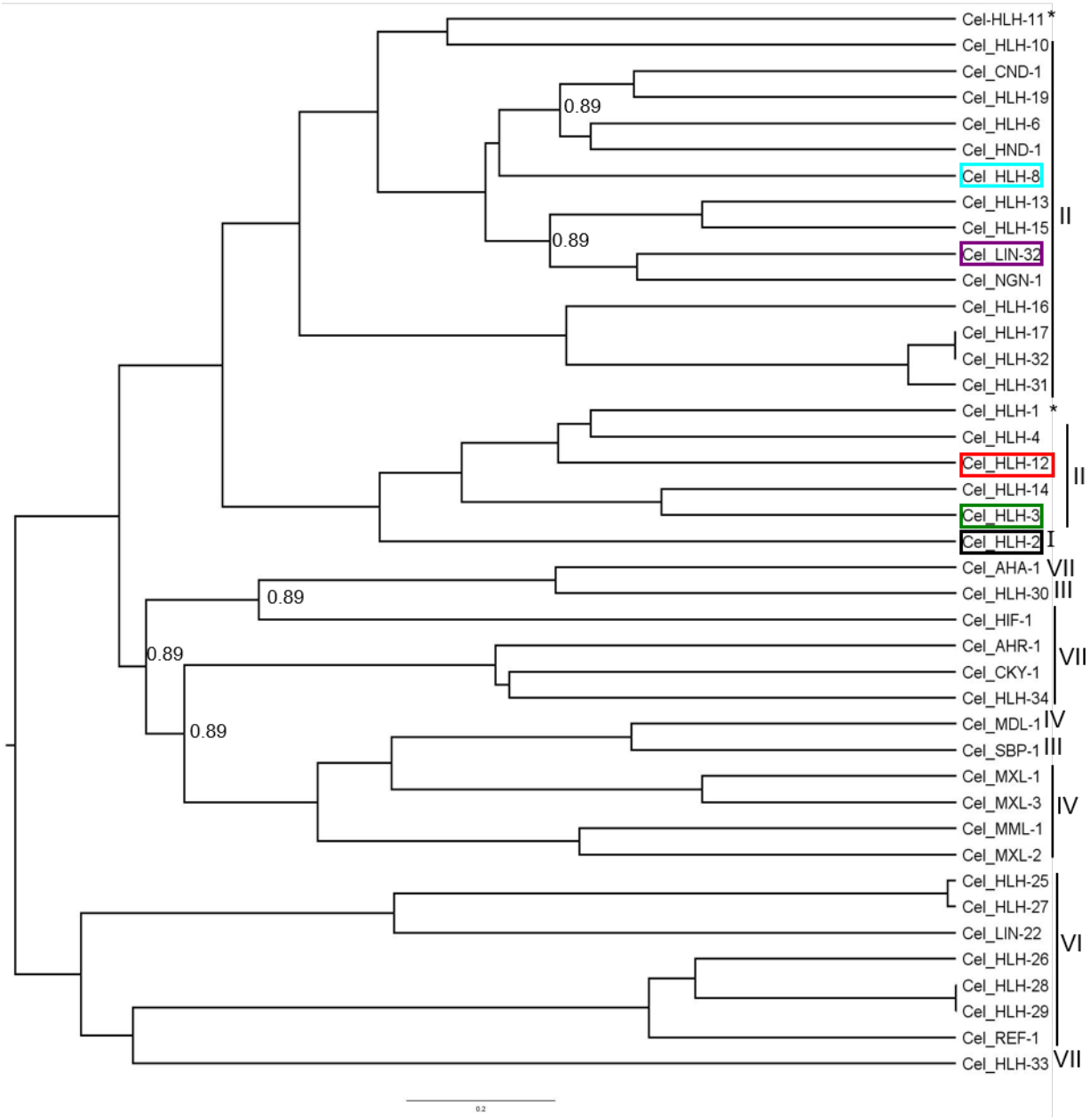


Figure 4.10. Phylogenetic analysis of all *C. elegans* bHLH genes. Only the bHLH domain was used for this analysis. Classes are indicated on right; * indicates orphan/unknown (see Table 4.3; Grove et al., 2009; Ledent and Vervoort, 2002). bHLH code genes are marked with colored boxes. Only posterior probabilities <0.9 are shown; branch lengths are in substitutions/site.

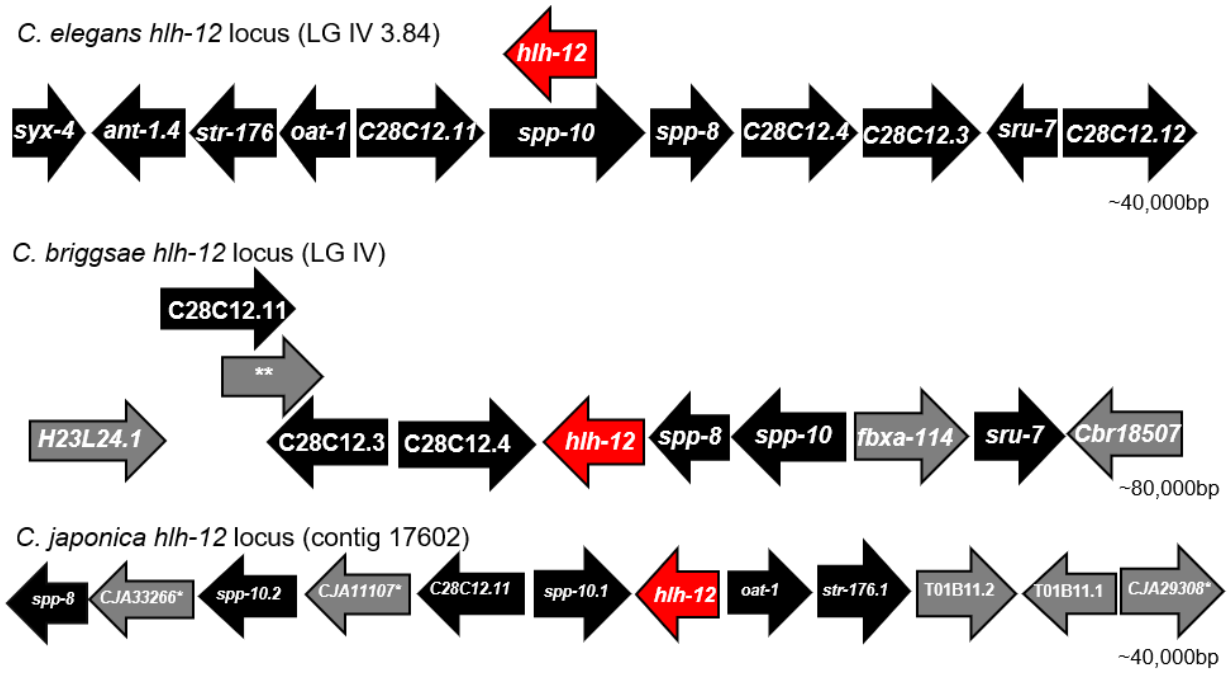


Figure 4.11. Synteny of *hlh-12* loci in *C. elegans, C. briggsae*, and *C. japonica*. Shown are the 10 most proximal genes to each *hlh-12* ortholog, including in *C. elegans spp-10* for a total of 11. Figure not to scale; approximate size of each locus shown on right. *hlh-12* is shown in red. In *C. briggsae* and *C. japonica*, black arrows indicate that gene is an ortholog of proximal gene in *C. elegans,* and *C. elegans* name is displayed. Grey arrows indicate that ortholog is not present at the *C. elegans hlh-12* locus. ** predicted tRNA.

Table 4.2. List of RNAi clones. All clones are from the Ahringer library (Kamath et al., 2003)

Gene	**Clone used**
ina-1	III-4N10
hlh-12	IV-4G20
pfd-1	IV-5N18
cdc-42	II-5P13

Table 4.3. List of annotated transcripts included in phylogenetic trees. Grey boxes indicate that no annotated ortholog exists.Gene families were assigned by consensus from Vervoort and Ledent 2002, Grove et al. 2009, and Wormbase version WS275; if there was no consensus, genes were marked as orphan. **P. pacificus* has a single annotated *hlh-3/hlh-14*. ** indicates that genes are predicted orthologs of E proteins (Kim et al., 2018). N/A indicates that gene was not included in analyses.

	bHLH	**Family**	***C. elegans***	***C. briggsae***	***C. japonica***	***C. remanei***	***P. pristionchus***
Class I	*hlh-2*	E	M05B5.5a.1	CBG21941.1	CJA10149.1	CRE10677.1	PPA01644.1 PPA33189.1
Class II	*hlh-3*	Achaete-Scute	T24B.6.1	CBG03260.1	CJA04859.1	CRE02308.1	PPA40124.1*
	hlh-4	Achaete-Scute	T05G5.2.1	CBG.10010.1	CJA10364.1	CRE25267.1	PPA05241.1
	hlh-6	Achaete-scute	T15H9.3.1	CBG06671.1	CJA24967a.1	CRE06984.1	PPA38786.1
	hlh-8	Twist	C02B8.4.1	CBG16905.1	CJA09725.1	CRE24053.1	PPA31506.1
	hlh-10	Atonal	ZK682.4.1	CBG19244.1	CJA14004.1	CRE19710.1	
	hlh-12	Orphan	C28C12.8.1	CBG18498.1	CJA32678.1	CRE03409.1	
	hlh-13	PTF1	F48D6.3.1	CBG14376.1	CJA18896.1	CRE00836.1	
	hlh-14	Achaete-Scute	C184A3.8.1	CBG02491.1	CJA06159.1	CRE11986.1	PPA40124.1*
	hlh-15	NSCL	C43H6.8.1	CBG02112.1	CJA03864.1	CRE07382.1	
	hlh-16	Beta3/Misti	DY3.3.1	CBG12364.1	CJA11661a.1	CRE29755.1	
	hlh-17	Olig**	F38C2.2.1	CBG13666.1	CJA47858	CRE15643-RA	
	hlh-19	Orphan	F57C12.3.1	CBG27394.1			
	hlh-31	Olig**	F38C2.8.1				
	hlh-32	Olig**	Y105C5B.29.1				
	cnd-1	NeuroD	C34210.7.2	CBG21175.1	CJA09215.1	CRE29369.1	
	hnd-1	Hand	C44C10.8.1	CBG15519.1	CJA21421.1	CRE05962.1	PPA30788.1
	lin-32	Atonal	T14F9.5.1	CBG14057.1	CJA03891a.1	CRE00941.1	PPA08416.1
	ngn-1	Neurogenin	Y69A2AR.29.1	CBG13881.1	CJA04915a.1	CRE20072.1	
Class VII	*hif-1*	Hif/Sim/Trh	F38A6.3a.1	CBG05540.1	CJA15222.1	N/A	PPA28926.1

Table 4.4. List of genome assemblies used in BLAST searches. Source abbreviations: WormBase ParaSite, PS; Caenorhabditis Genomes Project, CGP; European Nucleotide Archive, ENA.

Species	Strain	Genome assembly	BLAST type	Source
A. suum	Not given	PRJNA80881	BLAST+	PS
C. afra	JU1199	PRJEB11394	Online	CGP
C. angaria	PS1010	PJRNA51225	BLAST+	CGP
C. bovis	Not given	PJREB34497	Online	CGP/ENA
C. brenneri	PB2801	C brenneri-6.0.1b	Online	CGP
C. castelli	JU1956	PJREB11356	Online	CGP
C. doughertyi	JU1771	PJREB11002	Online	CGP
C. kamaaina	QG2077	QG2077 v1	Online	CGP
C. latens	PX534	PRJNA248912	BLAST+	CGP
C. monodelphis	JU1667	PRJEB7905	Online	CGP
C. nouraguensis	JU2079	PRJEB10884	Online	CGP
C. plicata	SB355	PRJEB11388	Online	CGP
C. sinica	JU1286	PRJEB11394	Online	CGP
C. tropicalis	JU1373	PRJNA53597	BLAST+	CGP
C. virilis	JU1968	PRJEB11359	Online	CGP
C. sp 21	NIC534	PRJEB12595	Online	CGP
C. sp 26	JU2190	PRJEB12596	Online	CGP
C. sp 28	QG2080	QG2080 v1	Online	CGP
C. sp 29	QG2083	QG2083 v1	Online	CGP
C. sp 31	JU2585	PRJEB12600	Online	CGP
C. sp 32	JU2788	PRJEB12601	Online	CGP
C. sp 34	NK74SC	NK74SC v710	Online	CGP
C. sp 38	JU2809	JU2809 v1	Online	CGP
C. sp 39	NIC564	NIC564 v1	Online	CGP
C. sp 40	JU2818	PRJEB7857	Online	CGP
H. bacteriophora	M31e	PRJNA13977	BLAST+	PS
M. belari	JU2817	JU2817 v2	Online	CGP
O. tipulae	CEW1	PRJEB15512	BLAST+	CGP
P. redivivus	PS2298	PRJNA186477	BLAST+	PS
P. oxycerus	Not given	Poikikolaimus oxycerus v1	Online	CGP

Table 4.5. BLAST search of *hlh-12* ortholog flanking regions. Identified hits are in bold.

Source	***C. elegans***	***C. briggsae***	***C. japonica***	***C. castelli***	***C. bovis***	***C. parvicauda***	***C. monodelphis***	*Notes*
***C.elegans* N**	only hit *hlh-12*	no hits	no hits	no hits	no hits	no hits	no hits	42bp fragment
***C. elegans* C**	only hit *hlh-12*	only hit *hlh-12*	only hit *hlh-12*	no hits	no hits	no hits	no hits	
***C. briggsae* N**	***set-13***	***sra-13***	no hits	no hits	no hits	no hits	no hits	
***C. briggsae* C**	***gly-14***	no hits	no hits	no hits	no hits	no hits	no hits	
***C. japonica* C**	only hit *hlh-12*	no hits	no hits	no hits	only hit *hlh-12*	no hits	no hits	*C. japonica hlh-12* bHLH domain begins at N terminal
***C. castelli* N**	no hits	no hits	no hits	only hit *hlh-12*	no hits	no hits	no hits	52bp fragment
***C. castelli* C**	no hits	***algn-7***	no hits	***linc-88***	no hits	no hits	no hits	

Table 4.6. Genomic loci of predicted *spp-10* and *hlh-12* orthologs. Size: top, kilobases(5' of referenced gene|3' of referenced gene); bottom, predicted coding sequences(5'|3'). For list of proximal genes, see Figure 11. * Scaffold contains only one gene.

	***spp-10* ortholog locus**				***hlh-12* ortholog locus**		
Species	Scaffold	Size	proximal genes present	*hlh-12* present?	Scaffold	Size	proximal genes present
C. uteleia	360	800(637|161) 256(205|50)	*str-176, C28C12.11, spp-8, C28C12.4, C28C12.3*	N	198	161(93|67) 43(21|21)	N
C. castelli	1649	625(180|459) 172(60|111)	*C28C12.4, C28C12.3*	N	1611	51(41|9) 15(11|3)	N
C. quiockensis	423	309(244|62) 94(78|15)	*spp-8, C28C12.4, C28C12.3*	N	157, 17266*	157: 147(28|118) 43(7|35)	N
C. virilis	2305	11(0|9) 3(0|2)	N	N	10	150(99|50) 27(19|7)	N
C. plicata	517	150(117|25) 20(17|2)	N	N	402	219(174|40) 24(20|3)	N

Supplemental Information

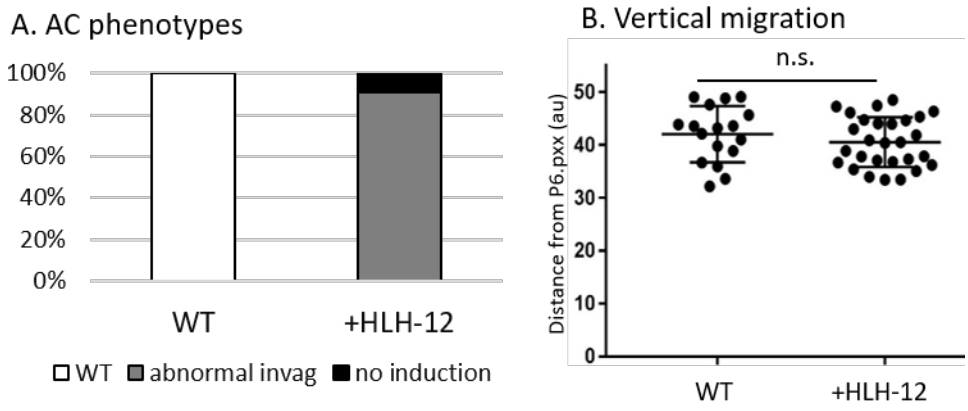


Figure S4.1. +HLH-12 AC phenotypes and vertical migration.

A. AC phenotypes seen in +HLH-12 ACs. "abnormal:" delayed invaginations, or invaginations with non-WT morphology. n = 20-23 for each genotype.

B. n.s. Kolmogorov-Smirnov, n = 18-25 for each genotype.

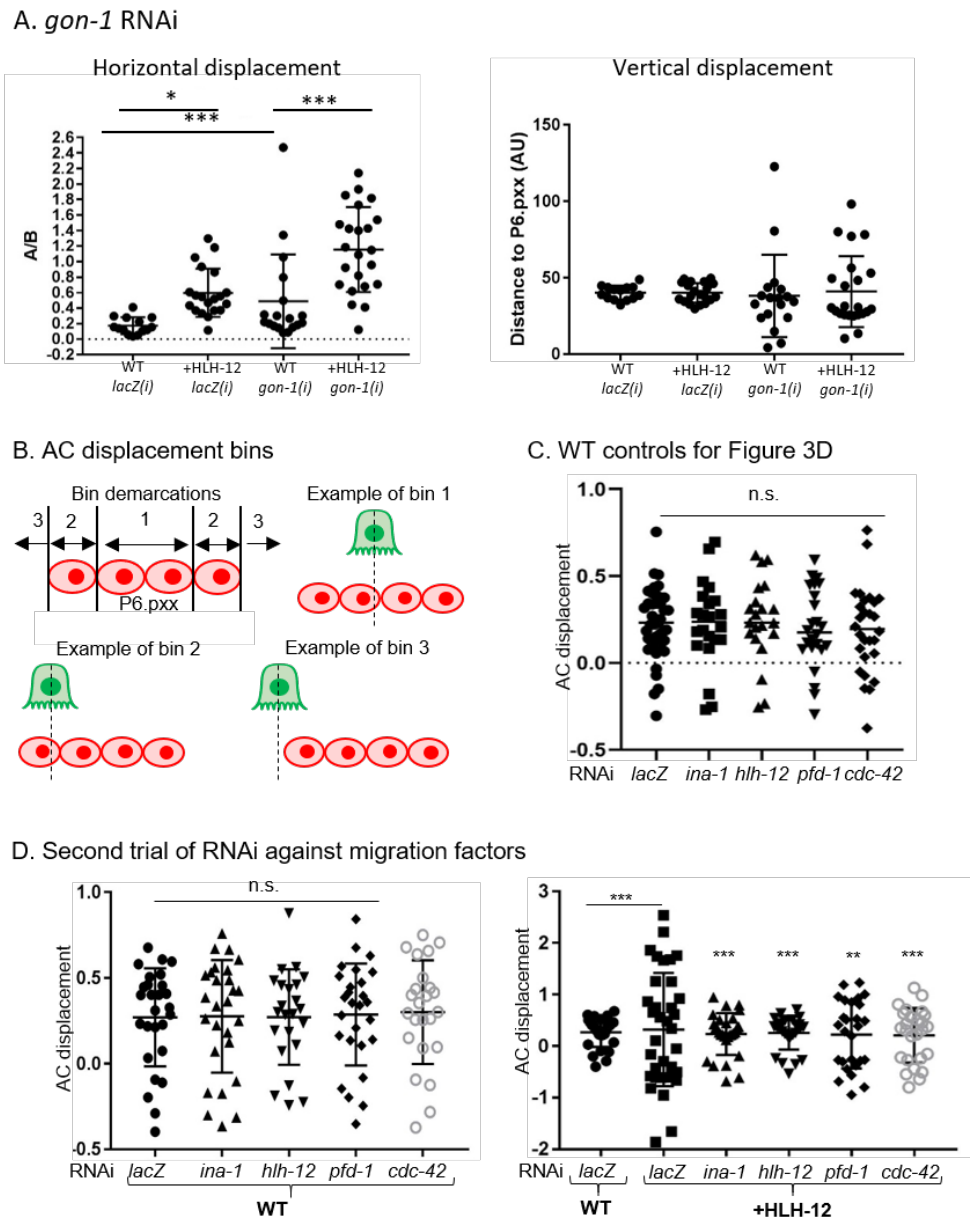


Figure S4.2. Further RNAi studies and scoring +HLH-12 AC displacement.

A. RNAi against *gon-1* causes overall gonad disruption that affects AC location in both WT and +HLH-12 animals. Note that horizontal displacement is quantified without regard to directionality. n = 20-25 for each genotype. Horizontal displacement: see Materials and Methods and Figure S2B for statistics used, ***p<0.001 compared to WT+*lacZ(RNAi)*. Vertical displacement: Kolmogorov-Smirnov, **p<0.01, ***p<0.001 compared to WT+*lacZ(RNAi)*.

B. Schematic of bins used to determine statistical significance of AC displacement. Animals were placed in bins according to the location of the center of the AC nucleus with respect to the P6.pxx nuclei. See Materials and Methods for more details.

C. WT controls of RNAi trial shown in Figure 2E. See Materials and Methods and Figure S2B for statistical details. n = 20-40, n.s. compared to WT+*lacZ(RNAi)*.

D. Second trial of RNAi against migration factors. See Materials and Methods and Figure S1C for statistical details. WT genotypes (left): n = 23-28, n.s. compared to WT+*lacZ(RNAi)*. +HLH-12 genotypes (right): n = 27-34, **p<0.01, ***p<0.001; when bar not shown, comparison is to +HLH-12+*lacZ(RNAi)*.

Figure S4.3. Phylogenetic tree of Class I and Class II bHLH proteins in 10 species. See attached file. The whole protein sequence was used for this analysis. bHLH code clades are highlighted: HLH-2, yellow; HLH-3/HLH-14, green; HLH-8, blue; HLH-12, red; LIN-32, purple. Numbers indicate posterior probabilities. Branch lengths are in substitutions/site.

Figure S4.4. Phylogenetic tree of Class I and Class II bHLH proteins in *Caenorhabditis* species. See attached file. bHLH code clades are highlighted: HLH-2, yellow; HLH-3/HLH-14, green; HLH-8, blue; HLH-12, red; LIN-32, purple. Numbers indicate posterior probabilities. Branch lengths are in substitutions/site.

Figure S4.5. Phylogenetic tree of Class I and Class II bHLH genes, and *hlh-1, hlh-11,* and *hif-1*, in a representative subset of *Caenorhabditis* species plus *Diploscapter* coronatus. See attached file. bHLH code clades are highlighted: HLH-2, yellow; HLH-3, green; HLH-8, blue; HLH-12, red; LIN-32, purple. Numbers indicate posterior probabilities. For species used, see Figure 7B. Branch lengths are in substitutions/site.

Table S4.1. Results of bHLH BLAST search in all nematode species. See attached file.

Table S4.2. Predicted orthologs of bHLH genes in all nematode species. See attached file.

Chapter 5: General Discussion

1. New insights into hDTC specification

1.1 *lin-32, hlh-2,* and *hlh-12* function at different steps in the lineage to promote hDTC fate and function

Based on the initial fosmid reporters discussed in Chapter 2, we found that LIN-32 was expressed starting in the specifying hDTC, and that it was expressed only in the hDTCs in the female somatic gonad (Figure 2.1, Figure S2.1). However, Sarah Finkelstein and I both saw expression of both N- and C-terminally-tagged endogenous lines starting in the hDTC parent stage; expression continued in both the hDTC and transiently and more dimly in the hDTC sister after the cell divided (Figure 3.4). *hlh-12* endogenous and fosmid expression both suggest that GFP::HLH-12 expression begins in the hDTC just after specification and continues until the cells stop migrating in L4, supporting *hlh-12's* known role in promoting hDTC migration (Figure 3.4, Figure S2.1, Tamai & Nishiwaki, 2007).

This expression of LIN-32 further up the hDTC lineage is interesting, as it fits with *lin-32's* known mechanisms of promoting fates in lineages such as the ray neurons (see Figure 1.6; Portman & Emmons, 2000). *lin-32's* expression pattern in the hDTC lineage is also reflective of *ceh-22b*/Nkx2.5/Tinman expression. In wild-type animals, *ceh-22b* is activated by the Wnt pathway; *ceh-22(lf)* animals specify 0 hDTCs, and overexpression under a heat-shock promoter converts the presumed hDTC sisters into migrating hDTCs (Lam et al., 2006). Based on the presence of *ceh-22* binding motifs in the *lin-32* 5' regulatory region and fosmid expression suggesting that *lin-32* was expressed in the hDTCs only and not their parents or sisters, Maria Sallee proposed that *ceh-22b* regulates *lin-32* expression in the hDTCs (M. Sallee, 2015). Our new understanding of *lin-32's* expression in the hDTC lineage strengthens this interpretation, and

further suggests that 1) *ceh-22b* could be regulating *lin-32*'s expression in the parent cell as well, and 2) loss of *ceh-22b* in the hDTC sister could explain why *lin-32* turns off in that cell shortly after its birth (see Figure 3.4).

Finally, while *hlh-12* is mostly thought of as a factor that promotes migration, our results suggest that *hlh-12* is at least partially necessary for hDTC fate as well. In the absence of both *hlh-12* and *lin-32*, the hDTCs failed to specify around 40% of the time, as indicated by a lack of both *lag-2p::rfp* and *hlh-2(fosmid)::gfp* expression (see Figure 2.2). The other half of the cells express both markers, however, suggesting that HLH-2 is forming homodimers that can promote hDTC fate; however, these *lag-2::rfp + hlh-2(fosmid)::gfp* cells do not migrate, suggesting that HLH-2:HLH-2 homodimers are insufficient for the migration function. In addition, we see germline proliferation even in animals lacking *lag-2p::rfp* and *hlh-2(fosmid)::gfp* expression in both erstwhile hDTCs (see Figure 2.2), suggesting that another pathway, such as Wnt acting through *ceh-22b/*Nkx2.5/Tinman, may be functionally redundant for the niche function (Lam et al., 2006).

In Chapter 2, we suggested a model in which the bHLH genes integrate positional, sex, lineage, and timing information to specify the proper cell fate (see Figure 2.4C). Based on our further understanding of *lin-32* expression in the hDTC lineage, I have refined this model to suggest that *lin-32* and *hlh-12* play distinct roles in hDTC specification, likely along with HLH-2 homodimers in each case (Figure 5.1). In this new model, positional information from the Wnt pathway is transmitted through *ceh-22b* expression in the hDTC parent, which turns on *lin-32*. After the cell divides, LIN-32:HLH-2 heterodimers and HLH-2 homodimers promote *hlh-2* expression, *lag-2* expression, and HLH-12 expression in the hDTC only, which results in a fully specified cell. Notably, this model explains why overexpression of HLH-2 from a ubiquitous heat-shock promoter can cause ectopic hDTCs to specify from their sisters (Karp & Greenwald, 2004):

as the sisters are born with LIN-32 inherited from their parent, the presence of HLH-2 allows for functional LIN-32 heterodimers and HLH-2 homodimers that promote full specification of hDTC fate.

1.2 Remaining questions about hDTC specification

While this model explains how the bHLH code genes can function together to promote hDTC specification, it leaves many questions unanswered. First, the model suggests that *ceh-22b* turns on *lin-32* expression, and that HLH-2 homodimers and possibly LIN-32:HLH-2 heterodimers turn on *hlh-12* expression, but we do not know what causes *hlh-2* expression. Preliminary data suggest that HLH-2 expression begins in the hDTC and is not present in the parent or its sister (Michelle Attner, personal communication); thus, HLH-2 expression appears to add single-cell resolution to a broader swathe of LIN-32 expression to cause precise cell fate specification. This is the opposite of the usual paradigm, in which Class I proteins are expressed broadly and tighter Class II expression determines which cell fates specify (Massari & Murre, 2000); HLH-2 has more restricted expression than Class I proteins in other organisms (Krause et al., 1997a), raising the question of whether the Class I and Class II roles have swapped in other cell fate specification events in *C. elegans.* In addition, this hypothesis that HLH-2 expression determines which cell becomes the hDTC makes the question of what turns on *hlh-2* even more important. *hlh-2's* restriction to the distal daughter implicates Wnt signaling, as does the overexpression experiment mentioned above (Lam et al., 2006): overexpression of CEH-22B in the sister cell might be enough to induce *hlh-2* expression there, which could allow for the formation of HLH-2 homodimers and HLH-2:LIN-32 heterodimers and promote hDTC cell fate. However, it might be difficult to tease apart any role that Wnt signaling plays in *hlh-2*'s expression from *ceh-22b*'s role in promoting *lin-*

32 expression. An option would be to search for CEH-22 consensus sites in the *hlh-2* promoter and then mutate them on a reporter transgene to assess expression in the hDTCs.

Another aspect of hDTC specification that this new model makes clear is the functional redundancy of HLH-2 homodimers and LIN-32 heterodimers. For example, *lin-32(0)* animals have a wild-type phenotype, suggesting that *hlh-12* has been turned on by HLH-2 homodimers; however, *hlh-12(0); lin-32(0)* hDTCs fail to specify around half the time, suggesting that in these cells the HLH-2 homodimers were insufficient for HLH-12 expression (Figure 2.2). The *hlh-12* 5' regulatory region contains both LIN-32-specific E boxes and those that can be bound by either LIN-32 heterodimers or HLH-2 homodimers, further supporting functional redundancy in this instance. In addition, the lack of *hlh-2* expression in some *lin-32(0); hlh-12(0)* hDTCs could be interpreted as evidence of an autoregulatory loop amongst the hDTC bHLH genes. Again, the fact that *hlh-2(fosmid)::gfp* is expressed in half of the *lin-32(0); hlh-12(0)* cells suggests that HLH-2 homodimer is sufficient for this autoregulation, but only around half the time. The ideal experiment to separate the two roles of heterodimer and homodimer would be to engineer an HLH-2 that can homodimerize but not heterodimerize; a proposed cysteine mutant exists that disrupts homodimerization, but it has yet to work in our hands (Maria Sallee, personal communication). The next-best thing would be to investigate E boxes on something like *lag-2*, which is a known HLH-2 homodimer target in the AC (Karp & Greenwald, 2003). Since LIN-32 heterodimers can bind to the reverse complement of HLH-2 homodimer E boxes, preventing mutational analysis (Grove et al., 2009), I could instead mutate the other LIN-32 heterodimer-specific E boxes and see if this affects *lag-2* expression in the hDTCs.

Finally, my pro-AC transformation results show that addition of HLH-12 and LIN-32 together to the pro-ACs results in *ceh-22b* expression with ~50% penetrance. Since I expect that

ceh-22b is functioning upstream of the bHLH code, more specifically *lin-32*, this means that there must be some sort of feedback mechanism from the bHLH genes onto their upstream factors; as both LIN-32 and HLH-12 expression are required for *ceh-22b* to turn on in the reprogrammed pro-ACs (see Figure 4.4; LIN-32 data not shown), it suggests that the feedback mechanism doesn't emanate from a single bHLH gene *per se*, but rather from a specified hDTC: either a downstream target of the bHLH genes, or a combination of LIN-32, HLH-12, and possibly HLH-2 dimers. Such feedback mechanisms may be necessary for cell fate or function maintenance in their endogenous setting.

2. Exploring the transience of pro-AC→DTC reprogramming

As discussed in Chapters 2 and 3, I am able to achieve both marker and morphological changes reminiscent of either mDTCs or hDTCs when I express their full bHLH code in the pro-ACs (Figure 2.4B, Figure 3.2). However, I see that this change in cell fate is transient, and in particular the morphological changes are lost by the L4 stage (Figure 3.2A,B). This suggests that while the cells initially change fate, there is some block to their achieving full and permanent reprogramming.

There are a few reasons why these changes in cell fate might be transient. Our first assumption was that the *hlh-2prox* promoter element, which expresses in the wild-type hermaphrodite proximal gonad only, was turned off as the cells became either mDTCs or hDTCs. However, I see constant expression of the fluorescent proteins from my ectopic expression constructs; while they do include the *sl2* splicing acceptor sequence, which may have some basal promoter activity (Justin Shaffer, personal communication), this suggests that *hlh-2prox* is unable to sense its changing cellular environment and thus that the cells are producing constant bHLH

protein even after transformation. Instead, I propose two alternative explanations for the loss of acquired cell fate over time in the pro-AC→DTC paradigms (Figure 5.2). The first is that using *hlh-2prox* results in bHLH proteins expressed at levels lower than in their endogenous contexts, meaning that *hlh-2prox* simply doesn't provide enough dimers of the appropriate types to achieve permanent reprogramming. The second is that there is a change in chromatin state in the AC beginning in L3, which might affect any reprogramming achieved before that point.

In the first hypothesis, high initial amounts of bHLH protein achieve a successful cell fate change; over time, as *hlh-2prox* expression lowers (M. D. Sallee & Greenwald, 2015), less bHLH protein is produced to maintain the transformation, and more HLH-2 is freed as a result to form HLH-2:HLH-2 homodimer and promote AC fate and functions. There are a few reasons for believing that *hlh-2prox* might express insufficient amounts of bHLH protein for a full cell fate conversion (Figure 5.2A). The first is that several of our endogenously-tagged bHLH alleles express very brightly in the wild-type regulatory cells, especially *hlh-12* in both the hDTCs and LC, and in contrast *hlh-2prox* is a relatively weak promoter (Maria Sallee, personal communication).The second is that I drove *hlh-2prox::hlh-12* from a single-copy integrated transgene, but *hlh-2prox::hlh-8* from a multi-copy array, and found a much higher initial penetrance of reprogramming in the pro-AC→mDTC context despite using the same *hlh-2prox::lin-32* transgene in both. Almost 100% of animals with the mDTC code expressed the mDTC marker, while only ~50% of animals with the hDTC code expressed the hDTC marker (Figure 2.4B, Figure 3.2A). In addition, I find that both HLH-8 and HLH-12 are stable in up to four cells in the proximal gonad, but I see transformation of four cells at a much higher penetrance in the pro-AC→mDTC context than in the pro-AC→hDTC context; as *hlh-2prox* expression is higher in the α cells than the β cells, the most likely explanation is that the higher levels of HLH-

8 overall are able to transform all four cells, while the lower levels of HLH-12 result in transformation of only the α cells in most cases.

Because of these findings, I initially tried to boost bHLH protein levels in the pro-ACs by using a system where Cre expression causes tissue-specific expression of proteins from a strong ubiquitous promoter (Justin Shaffer, personal communication); however, preliminary experiments suggested that the *hlh-2prox::cre* needed for this experiment was insufficient to enable strong bHLH protein expression in the pro-ACs. Another way to test this hypothesis would be to express *hlh-2prox::hlh-8* from a single-copy transgene integrated into the same LG1 site as the *hlh-2prox::hlh-12* (Pani and Goldstein, 2018). If the single-copy *prox::hlh-8* causes a lower penetrance of pro-AC→mDTC transformation in L2, it would support the hypothesis that bHLH protein levels are important in these transformations, and that adding even higher levels of bHLH protein might result in more permanent reprogramming. If the transformation penetrance remains higher than that seen with the pro-AC→hDTC context even with HLH-8 expression from a single-copy transgene, it would suggest that the pro-AC→mDTC transformation is easier to achieve. Notably, expression from our endogenously-tagged alleles suggest that both HLH-8 and LIN-32 are expressed only transiently in the mDTCs around the time of their specification, while HLH-12 is expressed in the hDTCs and LC throughout larval development (Figure 3.4; Sarah Finkelstein, personal communication); thus, it may be that mDTCs require less bHLH protein overall for their specification, suggesting that they may retain higher penetrance of transformation even when effected only by single-copy transgenes.

Another hypothesis for the transience of the pro-AC→DTC transformations is that a potential change in the wild-type AC chromatin state during L3 may block targets needed to maintain the cell fate change. During L2, the AC is specifying and then signaling to induce VPC

fate; during L3, the AC instead invades through the basement membrane (Sternberg, 2005). This change in AC function may be accompanied by a change in chromatin state, as the histone deacetylase *hda-1* is necessary for invasion (Matus et al., 2015). It is possible that the addition of bHLH genes to the pro-ACs results in transformation during L2, but that the chromatin state still changes during L3 and impedes access to the genes necessary for permanent reprogramming (Figure 5.2B). The easiest way to test this hypothesis would be to perform RNAi against *hda-1* in the pro-AC→mDTC or pro-AC→hDTC backgrounds and look for evidence of sustained reprogramming, such as DTC-like morphology retained in the late L3 stage. If *hda-1* RNAi treatment does sustain transformation, a full RNAi screen of chromatin modifiers might suggest if there is a larger-scale chromatin remodeling process happening in the L3 AC. Finally, if there is a change in chromatin state that occurs even in these reprogrammed cells, it suggests that the addition of bHLH genes alone is not sufficient to override all functions of the AC, and raises the possibility of parallel pathways controlling events in L3.

3. LC fate may be established by *hlh-3* and maintained by *lin-32*

3.1 *hlh-3* may act up the lineage to promote LC fate

As mentioned above, analysis of endogenously-tagged bHLH gene expression revealed that *lin-32* is expressed earlier in the hDTC lineage than its fosmid expression suggested. Another striking difference between the fosmid and CRISPR expression patterns is that I find HLH-3 expressed as early as the grandparent cells in both lineages with potential to become the LC (Figure 3.4). This recalls HLH-2's expression pattern in the AC lineage, and further supports the idea that *hlh-3* is the Class II bHLH most responsible for LC fate in the developing somatic primordium (see Figure 2.4A, Figure 3.4).

As HLH-3 requires HLH-2 to be functional, the next step is to find out if HLH-2 is also expressed in the LC parents or grandparents by looking at the CRISPR expression in the forming male somatic primordium. If HLH-2 is expressed, assuming that the HLH-2 is functional, this would suggest that HLH-2:HLH-3 dimers or a combination of hetero- and homodimers might function in specifying LC competence in the parent or grandparent cells, as HLH-2 homodimers do in the AC (Karp & Greenwald, 2004). Alternatively, if HLH-2 is not expressed in the LC lineage, that might suggest that the HLH-3 serves as a pool for faster LC specification once HLH-2 is present, as LIN-32 might do in hDTC specification.

3.2 *lin-32* may play a later role in LC fate maintenance or function

We found that a *lin-32* fosmid reporter expressed in the LC; however, our two endogenously-tagged lines did not show visible expression (Figure 3.4, Figure S3.2). This raised the important question of whether or not LIN-32 is required in the LC for its proper specification, as would be predicted from the bHLH code hypothesis. I found that *lin-32* is indeed required for LC specification, as *hlh-3(0); lin-32(0)* animals express the AC marker *cdh-3p::gfp* in their male proximal gonad cells (Figure 3.5). The enhancement of this phenotype by *hlh-12* RNAi suggests that the cells do retain some HLH-12 expression, though likely at low levels given the absence of *hlh-12* fosmid expression in *hlh-3(0)* "LCs" (Chapter 2). When I remove HLH-12 expression completely, this might have the effect of freeing up more HLH-2 protein to act as homodimers and promote aspects of AC fate, explaining the higher penetrance of AC marker expression in these animals.

I initially interpreted this genetic data as suggesting that LIN-32 is expressed in the future LC at the time of the forming somatic primordium, as loss of LIN-32 reprograms the fate that

would otherwise specify at this time, and as I have found that LIN-32 and HLH-8 when expressed at the primordium stage are capable of pro-AC→mDTC transformation, at least initially (Figure 2.4). However, the fosmid data show that LIN-32::GFP is expressed in the LC with low penetrance in L2 and high penetrance in L3, both after the LC is specified (Figures S2.1, S3.2B). As discussed earlier, differences in the transgene structures might cause different levels or expression patterns over time of LIN-32::GFP, and in particular the fosmid might not reflect true LIN-32 expression.

Another explanation for the timing of the fosmid expression, however, is that LIN-32 is not required for LC fate at the time of the somatic primordium, but instead is required for reinforcing some aspect of LC fate or function in L3, beyond the timeframe of our analysis. This later function of LIN-32 would also explain the timing of AC marker expression in the bHLH(-) LC experiments: I see the AC marker coming on in the male proximal gonad cells only in mid to late L3, a full larval stage later than its cognate expression in the AC (Figure 3.6A). In this situation, the cell would first become primed to be an AC by the absence of HLH-3 at the primordium stage, which enables AC potential (Figure 2.4); later in L3, the lack of LIN-32 would cause a full conversion to an AC (Figure 3.5, Figure 5.3). This interpretation is strengthened by the fact that I do not see any difference in phenotypes between the LCs of *hlh-2(bx108)* and *hlh-2(bx108); lin-32(0)* animals (Figure 3.6B). *hlh-2(bx108)* carries a mutation that interferes with dimerization; if LIN-32 did play some role in specification at the primordium stage, I would expect to see failure in specification resulting from the lessened amounts of functional dimers. Instead, the fact that these *hlh-2(bx108); lin-32(0)* LCs resemble *hlh-2(bx108)* LCs at least through the late L3 stage supports the hypothesis that LIN-32 is required at a later timepoint for some function of the LC or for maintenance of LC fate, and is not required during the somatic primordium stage (Figure 5.3).

This interpretation depends on the assumption that LIN-32 is acting in a cell-autonomous manner to promote LC fate; to confirm if this is the case, I would need to knock down LIN-32 in a more cell-specific manner. One method of cell-specific knockdown is to use the auxin-inducible degron system, which combines a degron tag with a cell-specific TIR1 for very precise knockdown (Zhang et al., 2015). This system is particularly attractive because the knockdown occurs only after the worms are plated on auxin, allowing for temporal control. I could combine a CRISPR-engineered LIN-32-degron allele with TIR1 driven by *nhr-67(2kb)p*, which I have found expresses in several cells in the male proximal gonad before LC specification and in the LC thereafter, or by *hlh-12(1080)p* for expression in the LC from just after specification through L4 (Tamai & Nishiwaki, 2007; Shai Shaham, personal communication; see Figure S3.1C). By adding auxin either at the time of primordium formation or later in L3, I could determine if LIN-32 does act in a cell-autonomous manner, and if so when it functions in the LC to further define its fate.

4. Hypotheses for later AC→LC transformation

While I saw pro-AC→DTC transformation beginning in the late L2 stage, and disappearing by the late L3 stage (Figures 2.4B, 2A,B), I saw the opposite trend when I added the LC bHLH code to the pro-ACs: only a single cell was transformed, and it did not achieve LC-like morphology until late L3 (Figure 3.3). This suggests both that the "pro-AC"→LC transformation is instead an AC→LC transformation, and that bHLH proteins function differently in effecting this AC→LC transformation than in the pro-AC→DTC transformations. While the issue of bHLH protein levels may still be relevant in this context, as I only see morphology changes in ~50% of animals, the fact that the transformation does appear to be occurring in L3 suggests that potential changes in

AC chromatin states happening at this time are not as important for this context. Here, I discuss several hypotheses for the delay in AC→LC transformation.

4.1 Hypothesis 1: occurrence of a LC/VD decision (Figure 5.4A)

The AC→LC reprogramming context is the only one in which the pro-ACs are being reprogrammed into a proximal-fate cell type. It is possible that there is some inbuilt mechanism preventing against improper specification of another proximal cell type, and that the bHLH proteins take longer to specify the LC as a result. Another clear difference between the LC and the DTCs is that the LC, like the AC, is specified as the result of LIN-12/Notch signaling between two cells; it is possible that the pro-ACs are reprogrammed into pro-LCs that then undergo an LC/VD decision. One way to test this would be to see if a VD-specific marker is expressed in the proximal female gonad in these animals, such as *egl-5*. As *egl-5* is expressed in most of the male gonad besides the LC and mDTCs (Kalis, Murphy, & Zarkower, 2010), its expression in an otherwise female somatic gonad would suggest that addition of the LC bHLH code to the pro-ACs resulted in an LC/VD decision in the hermaphrodite.

4.2 Hypothesis 2: timing of bHLH factors in WT specification might cause AC→LC delay (Figure 5.4B)

Another possibility for the delay in this transformation context is that bHLH factors act differently in specifying the WT LC. In particular, there is more evidence for activity further up the lineage. As mentioned earlier, *hlh-3* expression is first seen as early as the grandparents of the lineages with LC potential; *hlh-12* expression is seen just after the LC is specified, consistent with its role in promoting LC migration (see Figure 3.4, Tamai & Nishiwaki, 2007). At the moment,

we do not know if HLH-3 is functional in the LC parents or grandparents in part because we have no data on HLH-2 expression in those lineages. However, if HLH-3 is functional in the LC parents or grandparents, it might act as a pioneer factor and promote chromatin changes necessary for specification into an LC; if expressed only beginning in the eventual LC, it might take longer to effect these necessary changes and thus properly specify an LC. Examining HLH-2 expression in the pro-LC lineages would be the first step to determining if this hypothesis is correct. Interestingly, removal of HLH-3 alone is sufficient to cause AC competence in the erstwhile LC (Figure 2.4A), but I saw no changes in AC fate or function in the +HLH-3 animals (Figure S3.1A, B). If HLH-3 is indeed acting beginning in the parents or grandparents, it might require a cell division to specify an LC, meaning that removal of HLH-3 beginning in the grandparents is sufficient to convert the cells to pro-ACs, but that addition of HLH-3 to the pro-ACs is insufficient on its own to promote LC fate.

In addition to proper timing of HLH-3 for LC fate specification, the timing of LIN-32 expression might also result in a delayed reprogramming event in the pro-AC→LC context. Importantly, while I do not see LIN-32 expression in the LC with the CRISPR lines, my genetic data does support a model in which *hlh-3* acts first to specify LC fate in late L1, and then *lin-32* reinforces LC fate or function beginning in L3 (see Figures 5.3 and 5.4B). Because of this, pro-AC→LC transformation might happen in two stages, with *hlh-3* acting directly after *hlh-2prox* expression begins and *lin-32* acting later. As I see *lin-32* fosmid expression in the LC in L3, approximately two larval stages after the LC specifies in late L1, it might function in L4 in the pro-AC→LC context to promote full conversion to an LC fate. If the auxin experiments described above confirmed that LIN-32 is required cell autonomously in L3 for proper LC fate, it would lend

credence to this hypothesis; doing similar auxin experiments in this context would further strengthen the idea that full LC conversion happens after LIN-32 activity in later larval stages.

4.3 Hypothesis 3: HLH-3 or other bHLH code proteins may remain inactivated until after full AC specification (Figure 5.4C)

So far in this thesis, I have discussed Class II bHLH protein activity as depending solely on the presence of HLH-2. However, there is evidence that HLH-2 and Class II bHLH genes can be post-translationally modified to remain inactive even if their expression is stable. In particular, phosphorylation of a conserved serine residue in the second helix of the helix-loop-helix domain is predicted to interfere with DNA binding, thus inactivating even stably expressed protein (Quan et al., 2016). All of the bHLH code genes, including HLH-2, have a serine or threonine at this position, suggesting that all of them are capable of being inactivated in this way (M. Sallee, 2015).

At the moment, we have no indication of whether or not any of the bHLH code genes in any of the regulatory cells are targeted for inactivation in this manner. However, one hypothesis is this phosphorylation could be part of the AC/VU decision and help to downregulate HLH-2 activity in the specifying VUs, possibly downstream of LIN-12 signaling. Because the pre-AC also has some active LIN-12, this would mean that a portion of its bHLH proteins were also phosphorylated and thus inactivated as well. In particular, if HLH-3 protein was inactivated until after the AC was fully committed and LIN-12 signaling ceased, this could mean that the transformation is delayed simply because it could not start until after the AC specified. This relief of inactivation could also combine with one or both of the other hypotheses suggested above, creating an even longer delay in the pro-AC→LC transformation.

5. Conclusion

The *C. elegans* somatic gonad regulatory cells provide an elegant system for studying how cells specify their proper fates and functions. Discovery of the bHLH code made this system even more exciting for me, as it provided a clear and combinatorial mechanism for the specification and special features of each regulatory cell. However, upon further investigation, this code proved both more powerful and more complex than we imagined: while I can change regulatory cell fate using the code, and in general predict what the regulatory cell fate change will be, potential technical and endogenous blocks keep these changes from becoming more permanent. In addition, even in *C. elegans* the bHLH proteins function in different ways to promote the fate and functions of each regulatory cell type, and while I find that I can use the code as it works in *C. elegans* to explain the evolution of different gonad morphologies, a closer study of bHLH phylogenies instead raised questions about whether or not bHLH proteins are involved at all in the patterning the gonads of other species.

My work advances our understanding of how these initially similar cells adopt fates and functions very different from one another, but it leaves a number of questions unanswered. How do the bHLH proteins recognize and consolidate information about the cell's sex, location, lineage, and more? Are there parallel pathways in the AC that prevent bHLH-only-induced full and permanent reprogramming? Are there further blocks to reprogramming of cells like the hDTCs, which must make the fundamentally important germline, or are all regulatory cells equally susceptible to transformation? Are other nematode species also using versions of a bHLH code to pattern their gonads, or do they use completely different mechanisms? I am hopeful that future work will address some of these questions and further our knowledge of the bHLH code.

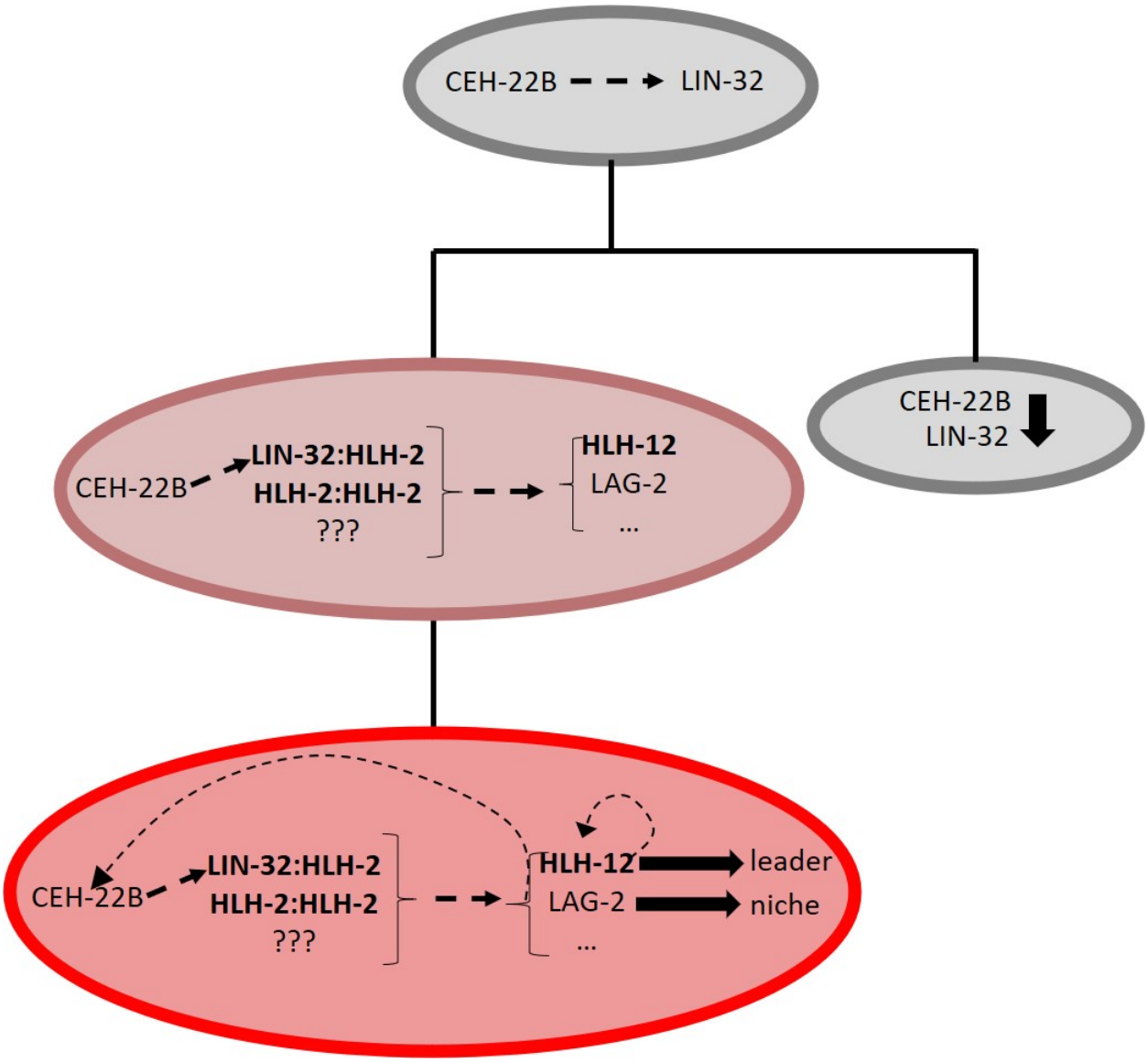


Figure 5.1. New model for hDTC specification. In the parent cell (top), CEH-22B turns on LIN-32 expression. After the cell divides, the sister cell (right) downregulates LIN-32 and possibly CEH-22B. In the specifying hDTC (left, top), LIN-32:HLH-2 and HLH-2:HLH-2 dimers, among other transcription factors, promote expression of HLH-12, LAG-2, and other targets. As the cell specifies (left, bottom), HLH-12 regulates its own expression and promotes the leader function, LAG-2 promotes the niche function, and either bHLH or downstream targets promote CEH-22B expression.

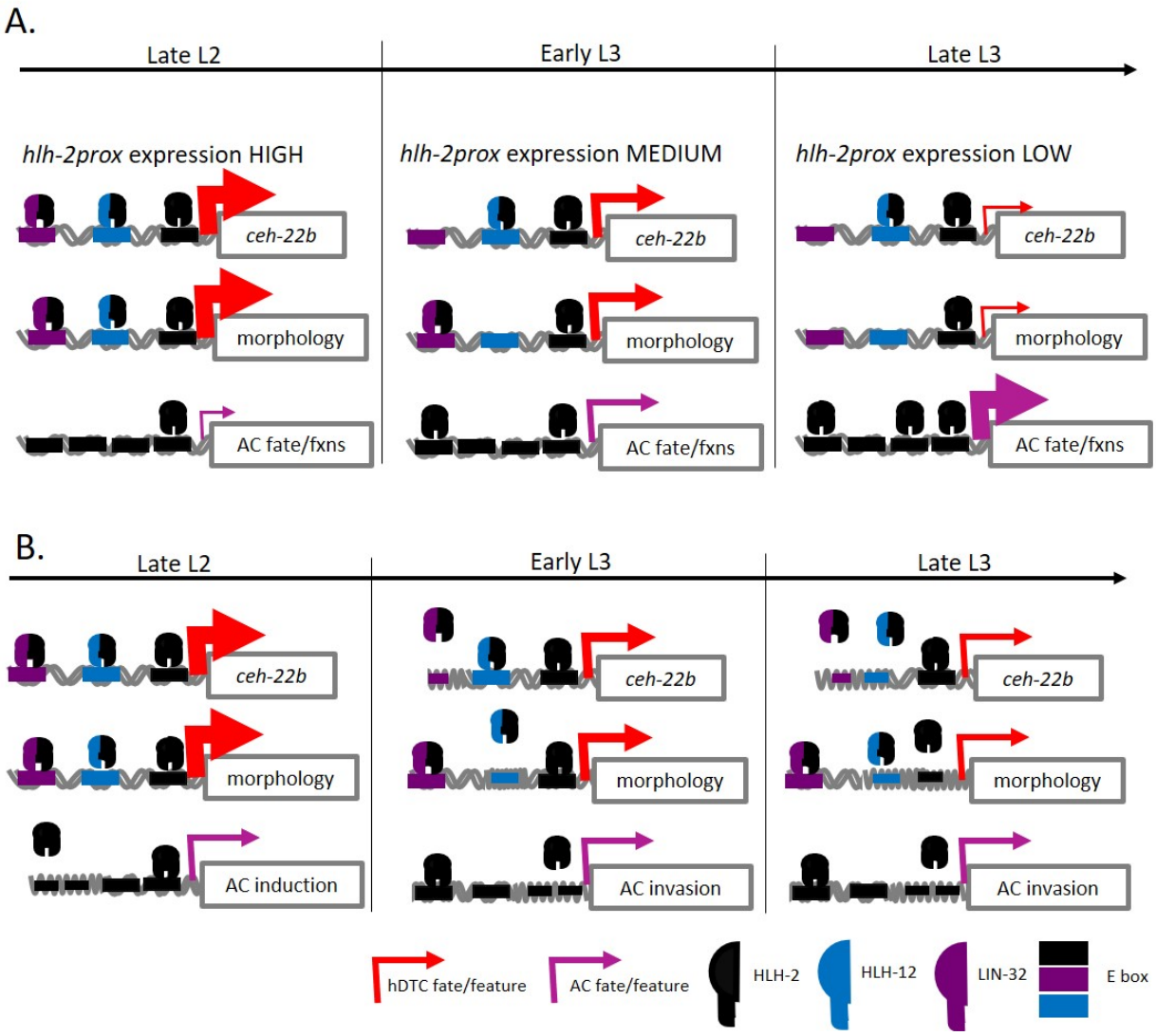


Figure 5.2. Alternate hypotheses for blocks to full pro-AC reprogramming. The pro-AC→hDTC transformation is shown, but the same can be implied for the pro-AC→mDTC context. Size of arrows reflects strength of expression.

A: A decrease in *hlh-2prox::bHLH* expression over time leads to less Class II bHLH protein in later stages and more available HLH-2 protein as a result, resulting in downregulation of hDTC-associated genes and functions and upregulation of AC-associated genes and functions.

B. Changes in AC chromatin state in L3 result in hDTC-associated targets being less available.

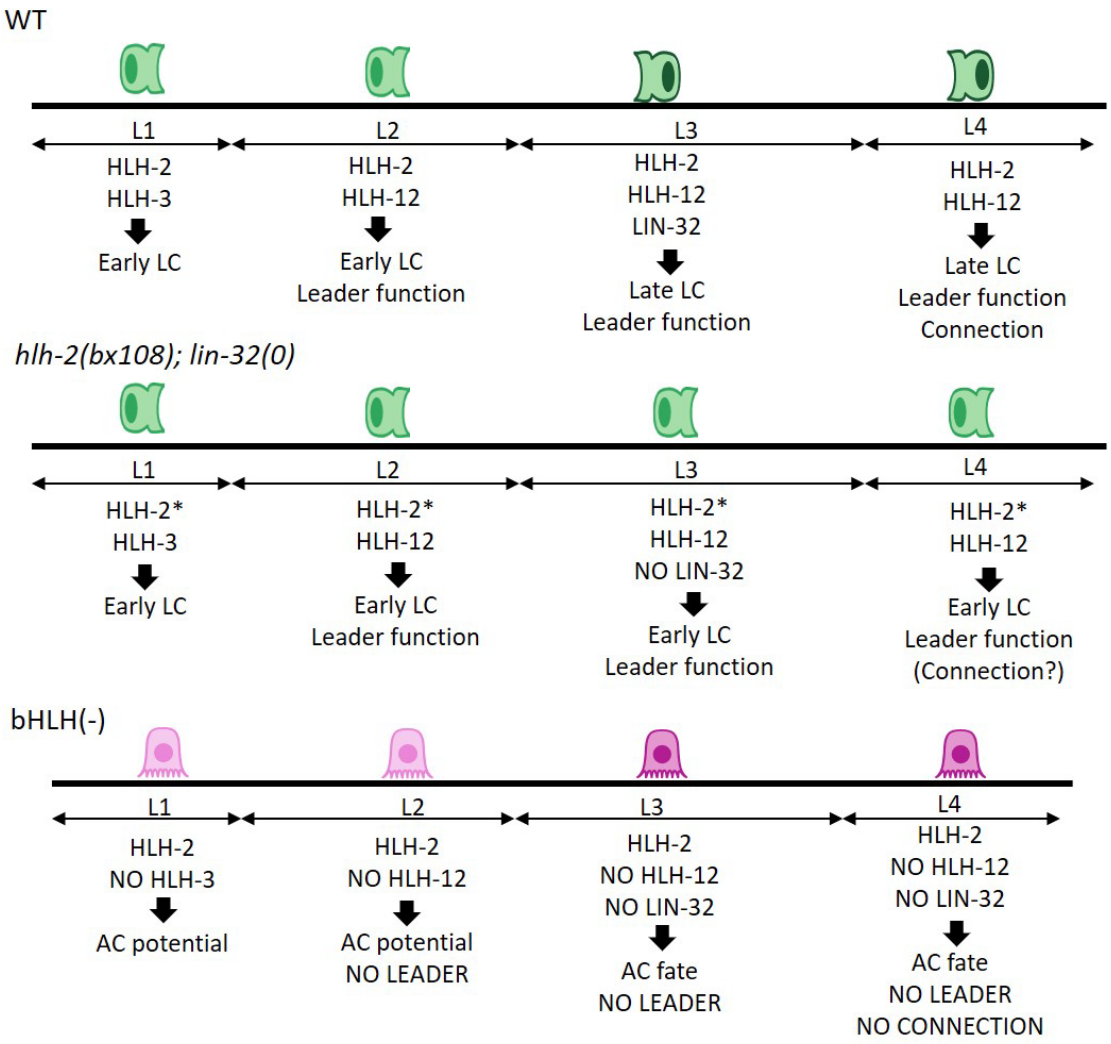


Figure 5.3. A potential model for LC specification. In the WT LC (top), HLH-2 and HLH-3 act in late L1 to promote "early LC" fate. In L2, HLH-12 promotes the leader function of the LC. In L3, LIN-32 promotes some further maintenance of fate or LC function to result in the "late LC," which continues migrating and connects with the cloaca in L4. The *hlh-2(bx108); lin-32(0)* LC (middle) specifies as an early LC due to available HLH-2 and HLH-3 in L1, and has present though impaired leader function in L2. In L3, lack of LIN-32 results in failure to specify as a "late LC," and a possible connection defect as a result. * indicates HLH-2 protein has compromised dimer ability. In the bHLH(-) LC (bottom), lack of HLH-3 in L1 results as the cell adopting AC potential, and lack of HLH-12 in the L2 results in a lack of leader function. Lack of LIN-32 in L3 results in the cell fully differentiating into an AC.

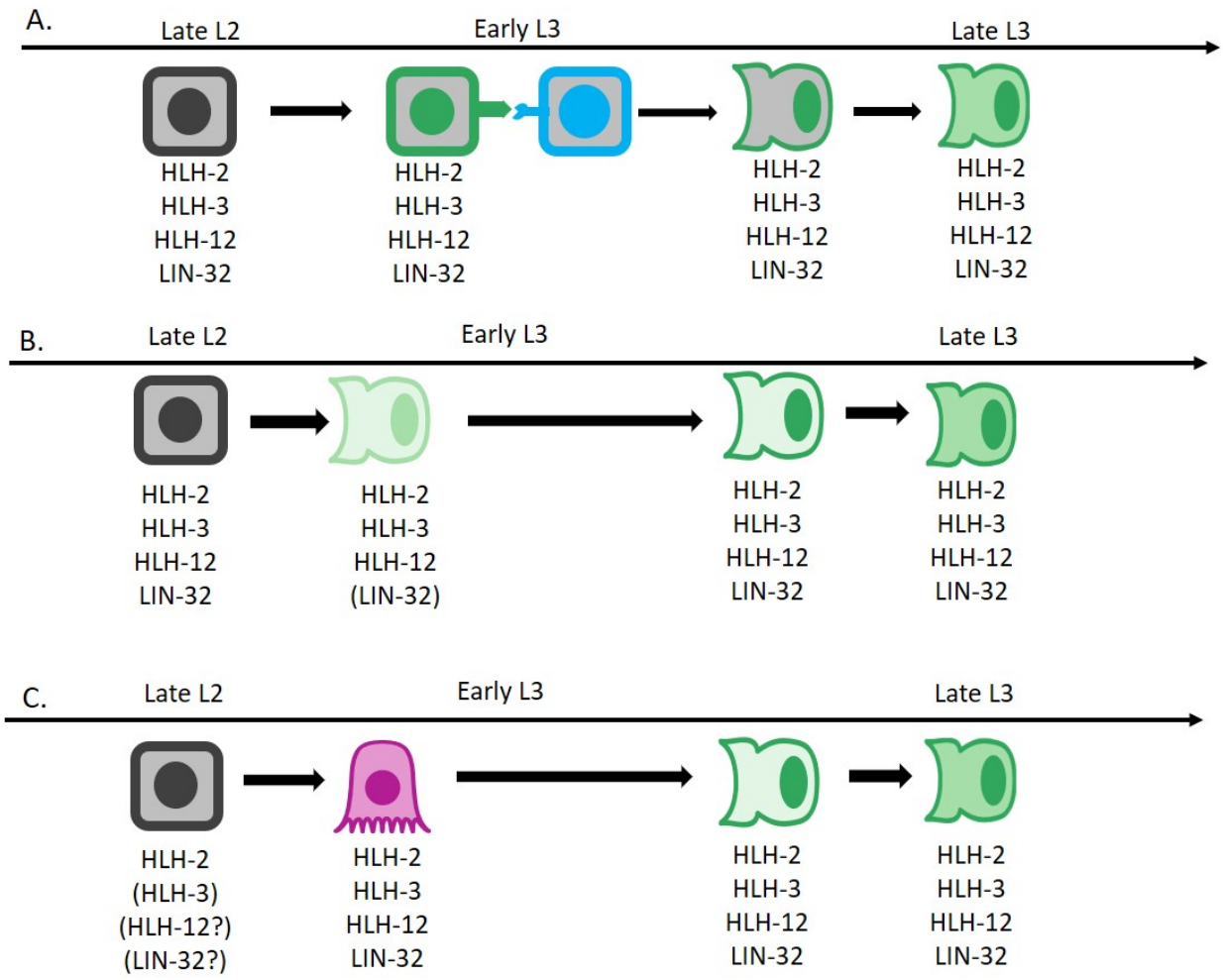


Figure 5.4. Alternate hypotheses for the delay in AC→LC transformation. The entire bHLH complement of the cells are shown at all times; proteins expected to be inactive at any stage are surrounded by parentheses.

A. Adding the LC code to the pro-ACs may result in cells first adopting a pre-LC fate and undergoing a LIN-12/Notch-mediated fate decision with the neighboring VU (shown in blue) before specifying fully as an LC and adopting LC-like morphology (dark green).

B. The LC-fated cell may first differentiate into an "early" LC (pale green, see Figure 3), until LIN-32 activity during L3 promotes its differentiation into a "late" LC with the expected morphology (dark green, see Figure 3). HLH-3 activity may also take longer than predicted to promote LC competence or fate (see section 4.2).

C. HLH-3 and other LC bHLH code proteins may be inactivated at first until the cell fully commits to AC fate, at which point they promote transformation to LC fate.

Works cited

Abraham, M. C., Lu, Y., & Shaham, S. (2007). A Morphologically Conserved Nonapoptotic Program Promotes Linker Cell Death in Caenorhabditis elegans. *Developmental Cell*, *12*(1), 73–86. https://doi.org/10.1016/j.devcel.2006.11.012

Alper, S., & Podbilewicz, B. (2008). Cell fusion in Caenorhabditis elegans. *Methods in Molecular Biology*. Humana Press. https://doi.org/10.1007/978-1-59745-250-2_4

Antebi, A., Culotti, J. G., & Hedgecock, E. M. (1998). daf-12 regulates developmental age and the dauer alternative in Caenorhabditis elegans. *Development*, *125*(7), 1191–1205.

Atchley, W. R., & Fitch, W. M. (1997). A natural classification of the basic helix-loop-helix class of transcription factors. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(10), 5172–5176. https://doi.org/10.1073/pnas.94.10.5172

Attner, M. A., Keil, W., Benavidez, J. M., & Greenwald, I. (2019). HLH-2/E2A Expression Links Stochastic and Deterministic Elements of a Cell Fate Decision during C. elegans Gonadogenesis. *Current Biology*, *29*(18), 3094-3100.e4. https://doi.org/10.1016/j.cub.2019.07.062

Austin, J., & Kimble, J. (1987). glp-1 Is required in the germ line for regulation of the decision between mitosis and meiosis in C. elegans. *Cell*, *51*(4), 589–599. https://doi.org/10.1016/0092-8674(87)90128-0

Bain, G., Engel, I., Robanus Maandag, E. C., te Riele, H. P., Voland, J. R., Sharp, L. L., … Murre, C. (1997). E2A deficiency leads to abnormalities in alphabeta T-cell development and to rapid development of T-cell lymphomas. *Molecular and Cellular Biology*, *17*(8), 4782–4791. https://doi.org/10.1128/mcb.17.8.4782

Baker, N. E., & Brown, N. L. (2018, May 1). All in the family: Proneural bHLH genes and neuronal diversity. *Development (Cambridge)*. Company of Biologists Ltd. https://doi.org/10.1242/dev.159426

Barkoulas, M., Vargas Velazquez, A. M., Peluffo, A. E., & Félix, M. A. (2016). Evolution of New cis-Regulatory Motifs Required for Cell-Specific Gene Expression in Caenorhabditis. *PLoS Genetics*, *12*(9), e1006278. https://doi.org/10.1371/journal.pgen.1006278

Blelloch, R., & Kimble, J. (1999). Control of organ shape by a secreted metalloprotease in the nematode Caenorhabditis elegans. *Nature*, *399*(June), 586–590. https://doi.org/10.1038/21196

Brenner, S. (1974). The Geneticas of Caenorhabdztzs elegans. *Genetics*, *77*(1), 71–94. Retrieved from

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1213120&tool=pmcentrez&rendertype=abstract

Brockes, J. P., & Kumar, A. (2002). Plasticity and reprogramming of differentiated cells in amphibian regeneration. *Nature Reviews Molecular Cell Biology*, *3*(8), 566–574. https://doi.org/10.1038/nrm881

Byrd, D. T., & Kimble, J. (2009, December 1). Scratching the niche that controls Caenorhabditis elegans germline stem cells. *Seminars in Cell and Developmental Biology*. Elsevier Ltd. https://doi.org/10.1016/j.semcdb.2009.09.005

Cabello, J., Neukomm, L. J., Günesdogan, U., Burkart, K., Charette, S. J., Lochnit, G., … Schnabel, R. (2010). The Wnt pathway controls cell death engulfment, spindle orientation, and migration through CED-10/Rac. *PLoS Biology*, *8*(2), 1000297. https://doi.org/10.1371/journal.pbio.1000297

Caudy, M., Vässin, H., Brand, M., Tuma, R., Jah, L. Y., & Jan, Y. N. (1988). daughterless, a Drosophila gene essential for both neurogenesis and sex determination, has sequence similarities to myc and the achaete-scute complex. *Cell*, *55*(6), 1061–1067. https://doi.org/10.1016/0092-8674(88)90250-4

Chang, W., Tilmann, C., Thoemke, K., Markussen, F.-H., Mathies, L. D., Kimble, J., & Zarkower, D. (2004). A forkhead protein controls sexual identity of the C. elegans male somatic gonad. *Development (Cambridge, England)*, *131*(6), 1425–1436. https://doi.org/10.1242/dev.01012

Chen, L., & Lopes, J. (2010). Multiple bHLH proteins regulate CIT2 expression in Saccharomyces cerevisiae. *Yeast*, *27*, 345–359. https://doi.org/10.1002/yea

Chen, N., & Greenwald, I. (2004). The lateral signal for LIN-12/Notch in C. elegans vulval development comprises redundant secreted and transmembrane DSL proteins. *Developmental Cell*, *6*(2), 183–192. https://doi.org/10.1016/S1534-5807(04)00021-8

Chesney, M. A., Lam, N., Morgan, D. E., Phillips, B. T., & Kimble, J. (2009). C. elegans HLH-2/E/Daughterless controls key regulatory cells during gonadogenesis. *Developmental Biology*, *331*(1), 14–25. https://doi.org/10.1016/j.ydbio.2009.04.015

Crews, S. T. (1998). Control of cell lineage-specific development and transcription by bHLH-PAS proteins. *Genes and Development*. https://doi.org/10.1101/gad.12.5.607

Crittenden, S. L., Lee, C. H., Mohanty, I., Battula, S., Knobel, K., & Kimble, J. (2019). Sexual dimorphism of niche architecture and regulation of the Caenorhabditis elegans germline stem cell pool. *Molecular Biology of the Cell*, *30*(14), 1757–1769. https://doi.org/10.1091/mbc.E19-03-0164

Cronmiller, C., & Cummings, C. A. (1993). The daughterless gene product in Drosophila is a

nuclear protein that is broadly expressed throughout the organism during development. *Mechanisms of Development*, *42*(3), 159–169. https://doi.org/10.1016/0925-4773(93)90005-I

Dejima, K., Hori, S., Iwata, S., Suehiro, Y., Yoshina, S., Motohashi, T., & Mitani, S. (2018). An Aneuploidy-Free and Structurally Defined Balancer Chromosome Toolkit for Caenorhabditis elegans. *Cell Reports*, *22*(1), 232–241. https://doi.org/10.1016/j.celrep.2017.12.024

Deng, T., Stempor, P., Appert, A., Daube, M., Ahringer, J., Hajnal, A., & Lattmann, E. (2020). The caenorhabditis elegans homolog of the evi1 proto-oncogene, egl-43, coordinates G1 cell cycle arrest with pro-invasive gene expression during anchor cell invasion. *PLoS Genetics*, *16*(3). https://doi.org/10.1371/journal.pgen.1008470

Dickinson, D. J., Pani, A. M., Heppert, J. K., Higgins, C. D., & Goldstein, B. (2015). Streamlined genome engineering with a self-excising drug selection cassette. *Genetics*, *200*(4), 1035–1049. https://doi.org/10.1534/genetics.115.178335

Doonan, R., Hatzold, J., Raut, S., Conradt, B., & Alfonso, A. (2008). HLH-3 is a C. elegans Achaete/Scute protein required for differentiation of the hermaphrodite-specific motor neurons. *Mechanisms of Development*, *125*(9–10), 883–893. https://doi.org/10.1016/j.mod.2008.06.002

Ellenberger, T., Fass, D., Arnaud, M., & Harrison, S. C. (1994). Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes and Development*, *8*(8), 970–980. https://doi.org/10.1101/gad.8.8.970

Emmons, S. W. (2005). Male development. *WormBook : The Online Review of C. Elegans Biology*. https://doi.org/10.1895/wormbook.1.33.1

Euling, S., & Ambros, V. (1996). Reversal of cell fate determination in Caenorhabditis elegans vulval development. *Development*, *122*(8), 2507–2515.

Farah, M. H., Olson, J. M., Sucic, H. B., Hume, R. I., Tapscott, S. J., & Turner, D. L. (2000). Generation of neurons by transient expression of neural bHLH proteins in mammalian cells. *Development*, *127*(4), 693–702.

Federation, A. J., Bradner, J. E., & Meissner, A. (2014). The use of small molecules in somatic-cell reprogramming. *Trends in Cell Biology*. https://doi.org/10.1016/j.tcb.2013.09.011

Félix, M. A., De Ley, P., Sommer, R. J., Frisse, L., Nadler, S. A., Thomas, W. K., … Sternberg, P. W. (2000). Evolution of vulva development in the Cephalobina (Nematoda). *Developmental Biology*, *221*(1), 68–86. https://doi.org/10.1006/dbio.2000.9665

Félix, M. A., & Sternberg, P. W. (1997). Two nested gonadal inductions of the vulva in nematodes. *Development*, *124*(1), 253–259.

Félix, M. a, & Sternberg, P. W. (1996). Symmetry breakage in the development of one-armed gonads in nematodes. *Development (Cambridge, England)*, *122*(7), 2129–2142.

Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B., & Burley, S. K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature*, *363*(6424), 38–45. https://doi.org/10.1038/363038a0

Fire, A., Harrison, S. W., & Dixon, D. (1990). A modular set of lacZ fusion vectors for studying gene expression in Caenorhabditis elegans. *Gene*, *93*(2), 189–198. https://doi.org/10.1016/0378-1119(90)90224-F

Frøkjær-Jensen, C., Davis, M. W., Sarov, M., Taylor, J., Flibotte, S., LaBella, M., … Jorgensen, E. M. (2014a). Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon. *Nature Methods*, *11*(5), 529–534. https://doi.org/10.1038/nmeth.2889

Frøkjær-Jensen, C., Davis, M. W., Sarov, M., Taylor, J., Flibotte, S., LaBella, M., … Jorgensen, E. M. (2014b). Random and targeted transgene insertion in Caenorhabditis elegans using a modified Mos1 transposon. *Nature Methods*, *11*(5), 529–534. https://doi.org/10.1038/nmeth.2889

Fukushige, T., & Krause, M. (2005). The myogenic potency of HLH-1 reveals wide-spread developmental plasticity in early C. elegans embryos. *Development*, *132*(8), 1795–1805. https://doi.org/10.1242/dev.01774

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Greenwald, I. S., Sternberg, P. W., & Robert Horvitz, H. (1983). The lin-12 locus specifies cell fates in caenorhabditis elegans. *Cell*, *34*(2), 435–444. https://doi.org/10.1016/0092-8674(83)90377-X

Grove, C. A., De Masi, F., Barrasa, M. I., Newburger, D. E., Alkema, M. J., Bulyk, M. L., & Walhout, A. J. M. (2009). A Multiparameter Network Reveals Extensive Divergence between C. elegans bHLH Transcription Factors. *Cell*, *138*(2), 314–327. https://doi.org/10.1016/j.cell.2009.04.058

Grove, C. A., De Masi, F., Inmaculada Barrasa, M., Newburger, D. E., Alkema, M. J., Bulyk, M. L., & Walhout, A. J. M. (n.d.-a). A Multiparameter Network Reveals Extensive Divergence between C. elegans bHLH Transcription Factors. *Cell*, *138*, 314–327. https://doi.org/10.1016/j.cell.2009.04.058

Grove, C. A., De Masi, F., Inmaculada Barrasa, M., Newburger, D. E., Alkema, M. J., Bulyk, M. L., & Walhout, A. J. M. (n.d.-b). *Supplemental Data A Multiparameter Network Reveals Extensive Divergence between C. elegans bHLH Transcription Factors SUPPLEMENTAL*

EXPERIMENTAL PROCEDURES Open Reading Frame Cloning. *Cell* (Vol. 138).

Gupta, B. P., Hanna-Rose, W., & Sternberg, P. W. (2012). Morphogenesis of the vulva and the vulval-uterine connection*. https://doi.org/10.1895/wormbook.1.152.1

Haag, E. S., Fitch, D. H. A., & Delattre, M. (2018). From "the worm" to "the worms" and back again: The evolutionary developmental biology of nematodes. *Genetics*, *210*(2), 397–433. https://doi.org/10.1534/genetics.118.300243

Haag, E. S., Helder, J., Mooijman, P. J. W., Yin, D., & Hu, S. (2018). The Evolution of Uniparental Reproduction in Rhabditina Nematodes: Phylogenetic Patterns, Developmental Causes, and Surprising Consequences. In *Transitions Between Sexual Systems* (pp. 99–122). Springer International Publishing. https://doi.org/10.1007/978-3-319-94139-4_4

Hagedorn, E. J., & Sherwood, D. R. (2011). Cell invasion through basement membrane: the anchor cell breaches the barrier This review comes from a themed issue on Cell-to-cell contact and extracellular matrix Edited by Arnoud Sonnenberg and Thomas Lecuit. *Current Opinion in Cell Biology*, *23*, 589–596. https://doi.org/10.1016/j.ceb.2011.05.002

Hall, D. H., Winfrey, V. P., Blaeuer, G., Hoffman, L. H., Furuta, T., Rose, K. L., … Greenstein, D. (1999). Ultrastructural features of the adult hermaphrodite gonad of Caenorhabditis elegans: Relations between the germ line and soma. *Developmental Biology*, *212*(1), 101–123. https://doi.org/10.1006/dbio.1999.9356

Harfe, B. D., Gomes, A. V., Kenyon, C., Liu, J., Krause, M., & Fire, A. (1998). Analysis of a Caenorhabditis elegans twist homolog identifies conserved and divergent aspects of mesodermal patterning. *Genes and Development*, *12*(16), 2623–2635. https://doi.org/10.1101/gad.12.16.2623

Hedgecock, E. M., Culotti, J. G., & Hall, D. H. (1990). The unc-5, unc-6, and unc-40 genes guide circumferential migrations of pioneer axons and mesodermal cells on the epidermis in C. elegans. *Neuron*, *4*(1), 61–85. https://doi.org/10.1016/0896-6273(90)90444-K

Henry, J. J., & Tsonis, P. A. (2010, November 1). Molecular and cellular aspects of amphibian lens regeneration. *Progress in Retinal and Eye Research*. Pergamon. https://doi.org/10.1016/j.preteyeres.2010.07.002

Hobert, O. (2002). PCR fusion-based approach to create reporter Gene constructs for expression analysis in transgenic C. elegans. *BioTechniques*, *32*(4), 728–730. https://doi.org/10.2144/02324bm01

Hwang, B. J., & Sternberg, P. W. (2004). A cell-specific enhancer that specifies lin-3 expression in the C. elegans anchor cell for vulval development. *Development*, *131*(1), 143–151. https://doi.org/10.1242/dev.00924

Jarriault, S., Schwab, Y., & Greenwald, I. (2008). *A Caenorhabditis elegans model for epithelial-*

neuronal transdifferentiation. Retrieved from www.pnas.org/cgi/content/full/

Kalis, A. K., Murphy, M. W., & Zarkower, D. (2010). EGL-5/ABD-B plays an instructive role in male cell fate determination in the C. elegans somatic gonad. *Developmental Biology*, *344*(2), 827–835. https://doi.org/10.1016/j.ydbio.2010.05.516

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., … Ahringer, J. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*, *421*(6920), 231–237. https://doi.org/10.1038/nature01278

Karp, X., & Greenwald, I. (2003). Post-transcriptional regulation of the E/Daughterless ortholog HLH-2, negative feedback, and birth order bias during the AC/VU decision in C. elegans. *Genes & Development*, *17*(24), 3100–3111. https://doi.org/10.1101/gad.1160803

Karp, X., & Greenwald, I. (2004). Multiple roles for the E/Daughterless ortholog HLH-2 during C. elegans gonadogenesis. *Developmental Biology*, *272*(2), 460–469. https://doi.org/10.1016/j.ydbio.2004.05.015

Kato, M., & Sternberg, P. W. (2009a). The C. elegans tailless/Tlx homolog nhr-67 regulates a stage-specific program of linker cell migration in male gonadogenesis. *Development*, *136*(23).

Kato, M., & Sternberg, P. W. (2009b). The C. elegans tailless/Tlx homolog nhr-67 regulates a stage-specific program of linker cell migration in male gonadogenesis. *Development*, *136*(23), 3907–3915. https://doi.org/10.1242/dev.035477

Kim, W., Underwood, R. S., Greenwald, I., & Shaye, D. D. (2018). Ortholist 2: A new comparative genomic analysis of human and caenorhabditis elegans genes. *Genetics*, *210*(2), 445–461. https://doi.org/10.1534/genetics.118.301307

Kimble, J., & Hirsh, D. (1979). The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans. *Developmental Biology*, *70*(2), 396–417. https://doi.org/10.1016/0012-1606(79)90035-6

Kinet, M. J., Malin, J. A., Abraham, M. C., Blum, E. S., Silverman, M. R., Lu, Y., & Shaham, S. (2016). HSF-1 activates the ubiquitin proteasome system to promote non-apoptotic developmental cell death in C. elegans. *ELife*, *5*(MARCH2016). https://doi.org/10.7554/eLife.12821

Kiontke, K., Barrière, A., Kolotuev, I., Podbilewicz, B., Sommer, R., Fitch, D. H. A., & Félix, M. A. (2007a). Trends, Stasis, and Drift in the Evolution of Nematode Vulva Development. *Current Biology*. https://doi.org/10.1016/j.cub.2007.10.061

Kiontke, K., Barrière, A., Kolotuev, I., Podbilewicz, B., Sommer, R., Fitch, D. H. A., & Félix, M. A. (2007b). Trends, Stasis, and Drift in the Evolution of Nematode Vulva Development. *Current Biology*, *17*(22), 1925–1937. https://doi.org/10.1016/j.cub.2007.10.061

Kodama, R., & Eguchi, G. (1995). From lens regeneration in the newt to in-vitro transdifferentiation of vertebrate pigmented epithelial cells. *Seminars in Cell and Developmental Biology*, *6*(3), 143–149. https://doi.org/10.1006/scel.1995.0020

Krause, M., Park, M., Zhang, J. M., Yuan, J., Harfe, B., Xu, S. Q., … Fire, A. (1997a). A C. Elegans E/Daughterless bHLH protein marks neuronal but not striated muscle development. *Development*, *124*(11), 2179–2189.

Krause, M., Park, M., Zhang, J., Yuan, J., Harfe, B., Xu, S., … Fire, A. (1997b). A C . elegans E / Daughterless bHLH protein marks neuronal but not striated muscle development, *2189*, 2179–2189.

Kutscher, L. M., Keil, W., & Shaham, S. (2018). RAB-35 and ARF-6 GTPases Mediate Engulfment and Clearance Following Linker Cell-Type Death. *Developmental Cell*, *47*(2), 222-238.e6. https://doi.org/10.1016/j.devcel.2018.08.015

Lam, N., Chesney, M. A., & Kimble, J. (2006). Wnt signaling and CEH-22/tinman/Nkx2.5 specify a stem cell niche in C. elegans. *Current Biology*, *16*(3), 287–295. https://doi.org/10.1016/j.cub.2005.12.015

Lambie, E. J., & Kimble, J. (1991). *Two homologous regulatory genes, lin-12 and glp-1, have overlapping functions*. *Development* (Vol. 112).

Ledent, Valérie, Paquet, O., & Vervoort, M. (2002). *Phylogenetic analysis of the human basic helix-loop-helix proteins*. Retrieved from http://genomebiology.com/2002/3/6/research/0030.1

Ledent, Valerie, & Vervoort, M. (2001). The Basic Helix-Loop-Helix Protein Family : Comparative Genomics and Phylogenetic Analysis Derivation of Comprehensive Sets. *Genome Biology*, *3*(6), 754–770. https://doi.org/10.1101/gr.177001.1

Li, J., & Greenwald, I. (2010). LIN-14 inhibition of LIN-12 contributes to precision and timing of C. elegans vulval fate patterning. *Current Biology*, *20*(20), 1875–1879. https://doi.org/10.1016/j.cub.2010.09.055

Liachko, N., Davidowitz, R., & Lee, S. S. (2009). Combined informatic and expression screen identifies the novel DAF-16 target HLH-13. *Developmental Biology*, *327*(1), 97–105. https://doi.org/10.1016/j.ydbio.2008.11.019

Lloret-Fernández, C., Maicas, M., Mora-Martínez, C., Artacho, A., Jimeno-Martín, Á., Chirivella, L., … Flames, N. (2018). A transcription factor collective defines the HSN serotonergic neuron regulatory landscape. *ELife*, *7*. https://doi.org/10.7554/eLife.32785

Lundin, V. F., Srayko, M., Hyman, A. A., & Leroux, M. R. (2008). Efficient chaperone-mediated tubulin biogenesis is essential for cell division and cell migration in C. elegans. *Developmental Biology*, *313*(1), 320–334. https://doi.org/10.1016/j.ydbio.2007.10.022

Luo, S., & Horvitz, H. R. (2017). The CDK8 Complex and Proneural Proteins Together Drive Neurogenesis from a Mesodermal Lineage. *Current Biology*, *27*(5), 661–672. https://doi.org/10.1016/j.cub.2017.01.056

Malin, J. Z., & Shaham, S. (2015). Cell Death in C. elegans Development. In *Current Topics in Developmental Biology* (Vol. 114, pp. 1–42). Academic Press Inc. https://doi.org/10.1016/bs.ctdb.2015.07.018

Mansfeld, J., Urban, N., Priebe, S., Groth, M., Frahm, C., Hartmann, N., … Ristow, M. (2015). Branched-chain amino acid catabolism is a conserved regulator of physiological ageing. *Nature Communications*, *6*(1), 1–12. https://doi.org/10.1038/ncomms10043

Marro, S., Pang, Z. P., Yang, N., Tsai, M. C., Qu, K., Chang, H. Y., … Wernig, M. (2011). Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell*, *9*(4), 374–382. https://doi.org/10.1016/j.stem.2011.09.002

Masoudi, N., Tavazoie, S., Glenwinkel, L., Ryu, L., Kim, K., & Hobert, O. (2018). Unconventional function of an Achaete-Scute homolog as a terminal selector of nociceptive neuron identity. *PLoS Biology*, *16*(4). https://doi.org/10.1371/journal.pbio.2004979

Massari, M. E., & Murre, C. (2000). MINIREVIEW Helix-Loop-Helix Proteins : Regulators of Transcription in Eucaryotic Organisms. *Molecular and Cellular Biology*, *20*(2), 429–440. https://doi.org/10.1128/MCB.20.2.429-440.2000.Updated

Matus, D. Q., Lohmer, L. L., Kelley, L. C., Schindler, A. J., Kohrman, A. Q., Barkoulas, M., … Sherwood, D. R. (2015). Invasive Cell Fate Requires G1 Cell-Cycle Arrest and Histone Deacetylase-Mediated Changes in Gene Expression. *Developmental Cell*, *35*(2), 162–174. https://doi.org/10.1016/J.DEVCEL.2015.10.002

Mayer, M. P. (1995). A new set of useful cloning and expression vectors derived from pBlueScript. *Gene*, *163*(1), 41–46. https://doi.org/10.1016/0378-1119(95)00389-N

Medwig-Kinney, T. N., Smith, J. J., Palmisano, N. J., Tank, S., Zhang, W., & Matus, D. Q. (2020). A developmental gene regulatory network for C. elegans anchor cell invasion. *Development (Cambridge)*, *147*(1). https://doi.org/10.1242/dev.185850

Meighan, C. M., Kann, A. P., & Egress, E. R. (2015). Transcription factor hlh-2/E/Daughterless drives expression of α integrin ina-1 during DTC migration in C. elegans. *Gene*, *568*(2), 220–226. https://doi.org/10.1016/j.gene.2015.05.030

Meighan, C. M., Kelly, V. E., Krahe, E. C., & Gaeta, A. J. (2015). α integrin cytoplasmic tails can rescue the loss of Rho-family GTPase signaling in the C.elegans somatic gonad. *Mechanisms of Development*, *136*, 111–122. https://doi.org/10.1016/j.mod.2014.12.006

Meighan, C. M., & Schwarzbauer, J. E. (2007). hermaphrodite gonad size and shape by vab-3 / Pax6-mediated regulation of integrin receptors, 1615–1620.

https://doi.org/10.1101/gad.1534807.by

Miller, R. M., & Portman, D. S. (2011). The wnt/β-catenin asymmetry pathway patterns the atonal ortholog lin-32 to diversify cell fate in a caenorhabditis elegans sensory lineage. *Journal of Neuroscience*, *31*(37), 13281–13291. https://doi.org/10.1523/JNEUROSCI.6504-10.2011

Morgenstern, B., & Atchley, W. R. (1999). Evolution of bHLH transcription factors: Modular evolution by domain shuffling? *Molecular Biology and Evolution*, *16*(12), 1654–1663. https://doi.org/10.1093/oxfordjournals.molbev.a026079

Motola, D. L., Cummins, C. L., Rottiers, V., Sharma, K. K., Li, T., Li, Y., … Mangelsdorf, D. J. (2006). Identification of Ligands for DAF-12 that Govern Dauer Formation and Reproduction in C. elegans. *Cell*, *124*(6), 1209–1223. https://doi.org/10.1016/j.cell.2006.01.037

Murre, C., Bain, G., van Dijk, M. A., Engel, I., Furnari, B. A., Massari, M. E., … Stuiver, M. H. (1994). Structure and function of helix-loop-helix proteins. *Biochemica et Biophysica Acta*, *1218*, 129–135. https://doi.org/10.1016/0005-2728(94)00175-5

Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., … Baltimore, D. (1989). Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell*, *58*(3), 537–544. https://doi.org/10.1016/0092-8674(89)90434-0

Nadarajan, S., Govindan, J. A., McGovern, M., Hubbard, E. J. A., & Greenstein, D. (2009). MSP and GLP-1/Notch signaling coordinately regulate actomyosin-dependent cytoplasmic streaming and oocyte growth in C. elegans. *Development (Cambridge, England)*, *136*(13), 2223–2234. https://doi.org/10.1242/dev.034603

Nakano, S., Ellis, R. E., & Horvitz, H. R. (2010). Otx-dependent expression of proneural bHLH genes establishes a neuronal bilateral asymmetry in C. elegans. *Development*, *137*(23), 4017–4027. https://doi.org/10.1242/dev.058834

Narasimhan, K., Lambert, S. A., Yang, A. W. H., Riddell, J., Mnaimneh, S., Zheng, H., … Hughes, T. R. (2015). Mapping and analysis of Caenorhabditis elegans transcription factor sequence specificities. *ELife*, *2015*(4), 1–53. https://doi.org/10.7554/eLife.06967

Newman, A. P., & Sternberg, P. W. (1996). Coordinated morphogenesis of epithelia during development of the Caenorhabditis elegans uterine-vulval connection. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(18), 9329–9333. https://doi.org/10.1073/pnas.93.18.9329

Newman, A. P., White, J. G., & Sternberg, P. W. (1995). The Caenorhabditis elegans lin-12 gene mediates induction of ventral uterine specialization by the anchor cell. *Development*, *121*, 263–271.

Portman, D. S., & Emmons, S. W. (2000a). The basic helix-loop-helix transcription factors LIN-32 and HLH-2 function together in multiple steps of a C. elegans neuronal sublineage. *Development (Cambridge, England)*, *127*(24), 5415–5426. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11076762

Portman, D. S., & Emmons, S. W. (2000b). The basic helix-loop-helix transcription factors LIN-32 and HLH-2 function together in multiple steps of a C. elegans neuronal sublineage. *Development*, *127*(24), 5415–5426. Retrieved from www.wormbase.org.

Quan, X. J., Yuan, L., Tiberi, L., Claeys, A., De Geest, N., Yan, J., … Hassan, B. A. (2016). Post-translational Control of the Temporal Dynamics of Transcription Factor Activity Regulates Neurogenesis. *Cell*, *164*(3), 460–475. https://doi.org/10.1016/j.cell.2015.12.048

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Rao, M. S., Dwivedi, R. S., Subborao, V., Usman, M. I., Scarpelli, D. G., Nemali, M. R., … Reddy, J. K. (1988). Almost total conversion of pancreas to liver in the adult rat: A reliable model to study transdifferentiation. *Biochemical and Biophysical Research Communications*, *156*(1), 131–136. https://doi.org/10.1016/S0006-291X(88)80814-3

Richard, J. P., Zuryn, S., Fischer, N., Pavet, V., Vaucamps, N., & Jarriault, S. (2011). Direct in vivo cellular reprogramming involves transition through discrete, non-pluripotent steps. *Development*, *138*(8), 1483–1492. https://doi.org/10.1242/dev.063115

Riddle, M. R., Spickard, E. A., Jevince, A., Nguyen, K. C. Q., Hall, D. H., Joshi, P. M., & Rothman, J. H. (2016). Transorganogenesis and transdifferentiation in C. elegans are dependent on differentiated cell identity. *Developmental Biology*, *420*(1), 136–147. https://doi.org/10.1016/j.ydbio.2016.09.020

Rödelsperger, C., Meyer, J. M., Prabh, N., Lanz, C., Bemm, F., & Sommer, R. J. (2017). Single-Molecule Sequencing Reveals the Chromosome-Scale Genomic Architecture of the Nematode Model Organism Pristionchus pacificus. *Cell Reports*, *21*(3), 834–844. https://doi.org/10.1016/j.celrep.2017.09.077

Rothman, J., & Jarriault, S. (2019). Developmental plasticity and cellular reprogramming in caenorhabditis elegans. *Genetics*, *213*(3), 723–757. https://doi.org/10.1534/genetics.119.302333

Sallee, M. (2015). Function and regulation of the transcription factor HLH-2 / E2A during gonadogenesis in Caenorhabditis elegans . *Thesis*. https://doi.org/10.7916/D82N511Z

Sallee, M. D., & Greenwald, I. (2015). Dimerization-driven degradation of C. elegans and human E proteins. *Genes & Development*, *29*(13), 1356–1361. https://doi.org/10.1101/gad.261917.115

Sallee, M. D., Littleford, H. E., & Greenwald, I. (2017). A bHLH Code for Sexually Dimorphic Form and Function of the C. elegans Somatic Gonad. *Current Biology*, *27*(12), 1853-1860.e5. https://doi.org/10.1016/j.cub.2017.05.059

Sapir, A., Choi, J., Leikina, E., Avinoam, O., Valansi, C., Chernomordik, L. V., … Podbilewicz, B. (2007). AFF-1, a FOS-1-Regulated Fusogen, Mediates Fusion of the Anchor Cell in C. elegans. *Developmental Cell*, *12*(5), 683–698. https://doi.org/10.1016/J.DEVCEL.2007.03.003

Schindler, A. J., & Sherwood, D. R. (2011). The transcription factor HLH-2/E/Daughterless regulates anchor cell invasion across basement membrane in C. elegans. *Developmental Biology*, *357*(2), 380–391. https://doi.org/10.1016/j.ydbio.2011.07.012

Schulze, J., & Schierenberg, E. (2011). Evolution of embryonic development in nematodes. *EvoDevo*, *2*(1), 1–17. https://doi.org/10.1186/2041-9139-2-18

Schwarz, E. M., Kato, M., & Sternberg, P. W. (2012). Functional transcriptomics of a migrating cell in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(40), 16246–16251. https://doi.org/10.1073/pnas.1203045109

Seydoux, G., & Greenwald, I. (1999). *Cell Autonomy of /h-12 Function in a Cell Fate Decision in C. elegans*. *Cell* (Vol. 57).

Seydoux, G., Savage, C., & Greenwald, I. (1993). Isolation and characterization of mutations causing abnormal eversion of the vulva in caenorhabditis elegans. *Developmental Biology*, *157*(2), 423–436. https://doi.org/10.1006/dbio.1993.1146

Sherwood, D. R., Butler, J. A., Kramer, J. M., & Sternberg, P. W. (2005). FOS-1 promotes basement-membrane removal during anchor-cell invasion in C. elegans. *Cell*, *121*(6), 951–962. https://doi.org/10.1016/j.cell.2005.03.031

Sherwood, D. R., & Plastino, J. (2018). Invading, Leading and Navigating Cells in Caenorhabditis elegans: Insights into Cell Movement in Vivo. *Genetics*, *208*, 53–78. https://doi.org/10.1534/genetics.117.300082

Sherwood, D. R., & Sternberg, P. W. (2003). Anchor cell invasion into the vulval epithelium in C. elegans. *Developmental Cell*, *5*(1), 21–31. https://doi.org/10.1016/S1534-5807(03)00168-0

Slos, D., Sudhaus, W., Stevens, L., Bert, W., & Blaxter, M. (2017). Caenorhabditis monodelphis sp. n.: Defining the stem morphology and genomics of the genus Caenorhabditis. *BMC Zoology*, *2*(1), 4. https://doi.org/10.1186/s40850-017-0013-2

Sommer, R., & Sternberg, P. (1994). Changes of induction and competence during the evolution of vulva development in nematodes. *Science*, *265*(5168). Retrieved from http://science.sciencemag.org/content/265/5168/114.long

Sternberg, P. W. (2005). Vulval development. *WormBook : The Online Review of C. Elegans Biology*, 1–28. https://doi.org/10.1895/wormbook.1.6.1

Stevens, J. D., Roalson, E. H., & Skinner, M. K. (2008). Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: Genomic approach to cellular differentiation. *Differentiation*, *76*(9), 1006–1022. https://doi.org/10.1111/j.1432-0436.2008.00285.x

Stoldt, V. R., Sonneborn, A., Leuker, C. E., & Ernst, J. F. (1997). Efg1p, an essential regulator of morphogenesis of the human pathogen Candida albicans, is a member of a conserved class of bHLH proteins regulating morphogenetic processes in fungi. *EMBO Journal*, *16*, 1982–1991.

Su, M. W., Merz, D. C., Killeen, M. T., Zhou, Y., Zheng, H., Kramer, J. M., … Culotti, J. G. (2000). Regulation of the UNC-5 netrin receptor initiates the first reorientation of migrating distal tip cells in Caenorhabditis elegans. *Development*, *127*(3), 585–594.

Sudhaus, W., & Hooper, D. J. (1994). Rhabditis (Oscheius) guentheri1 sp.n., an unusual species with reduced posterior ovary, with observations on the dolichura and insectivora groups (Nematoda: Rhabditidae). *Nematologica*, *40*(1–4), 508–533. https://doi.org/10.1163/003525994X00391

Sulston, J. E., & Horvitz, H. R. (1977). *Post-embryonic Cell Lineages of the Nematode, Caenorhabditis elegans*. *DEVELOPMENTAL BIOLOGY* (Vol. 56).

Sun, H., Ghaffari, S., & Taneja, R. (2007). bHLH-Orange Transcription Factors in Development and Cancer. *Translational Oncogenomics*, *2*, 107–120. https://doi.org/10.4137/tog.s436

Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Tamai, K. K., & Nishiwaki, K. (2007a). bHLH transcription factors regulate organ morphogenesis via activation of an ADAMTS protease in C. elegans. *Developmental Biology*, *308*(2), 562–571. https://doi.org/10.1016/j.ydbio.2007.05.024

Tamai, K. K., & Nishiwaki, K. (2007b). bHLH transcription factors regulate organ morphogenesis via activation of an ADAMTS protease in C. elegans. *Developmental Biology*, *308*(2), 562–571. https://doi.org/10.1016/j.ydbio.2007.05.024

Verghese, E., Schocken, J., Jacob, S., Wimer, A. M., Royce, R., Nesmith, J. E., … Wightman, B. (2011). The tailless ortholog nhr-67 functions in the development of the C. elegans ventral uterus. *Developmental Biology*, *356*(2), 516–528. https://doi.org/10.1016/j.ydbio.2011.06.007

Ward, J. D. (2014). Rapid and precise engineering of the caenorhabditis elegans genome with

lethal mutation co-conversion and inactivation of NHEJ repair. *Genetics*, *199*(2), 363–377. https://doi.org/10.1534/genetics.114.172361

Wei, X., Potter, C. J., Luo, L., & Shen, K. (2012). Controlling gene expression with the Q repressible binary expression system in Caenorhabditis elegans. *Nature Methods*, *9*(4), 391–395. https://doi.org/10.1038/nmeth.1929

Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B., & Miller, A. D. (1989). *Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD (muscle regulatory gene/MyoD retrovirus)*. *Developmental Biology* (Vol. 86).

Witte, H., Moreno, E., Rödelsperger, C., Kim, J., Kim, J. S., Streit, A., & Sommer, R. J. (2014). Gene inactivation using the CRISPR/Cas9 systemin the nematode Pristionchus pacificus. *Development Genes and Evolution*, *225*(1), 55–62. https://doi.org/10.1007/s00427-014-0486-8

Wong, M.-C., & Schwarzbauer, J. E. (2012). Gonad morphogenesis and distal tip cell migration in the *Caenorhabditis elegans* hermaphrodite. *Wiley Interdisciplinary Reviews: Developmental Biology*, *1*(4), 519–531. https://doi.org/10.1002/wdev.45

Zarkower, D. (2006). Somatic sex determination. *WormBook : The Online Review of C. Elegans Biology*. https://doi.org/10.1895/wormbook.1.84.1

Zhang, L., Ward, J. D., Cheng, Z., & Dernburg, A. F. (2015). The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. *Development (Cambridge, England)*, *142*(24), 4374–4384. https://doi.org/10.1242/dev.129635

Zhang, Y., Yu, Z., Fu, X., & Liang, C. (2002). Noc3p, a bHLH protein, plays an integral role in the initiation of DNA replication in budding yeast. *Cell*, *109*(7), 849–860. https://doi.org/10.1016/S0092-8674(02)00805-X

Zhuang, Y., Barndt, R. J., Pan, L., Kelley, R., & Dai, M. (1998). Functional Replacement of the Mouse E2A Gene with a Human HEB cDNA. *Molecular and Cellular Biology*, *18*(6), 3340–3349. https://doi.org/10.1128/mcb.18.6.3340

Zhao, C., Emmons, S.W. (1995). A transcription factor controlling development of peripheral sense organs in *C. elegans*. Nature *373*, 74–78.

Zhang, L., Ward, J. D., Cheng, Z., & Dernburg, A. F. (2015). The auxin-inducible degradation (AID) system enables versatile conditional protein depletion in C. elegans. *Development (Cambridge, England)*, *142*(24), 4374–4384. https://doi.org/10.1242/dev.129635

Zhang, Y., Yu, Z., Fu, X., & Liang, C. (2002). Noc3p, a bHLH protein, plays an integral role in the initiation of DNA replication in budding yeast. *Cell*, *109*(7), 849–860. https://doi.org/10.1016/S0092-8674(02)00805-X

Zhu, Z., Liu, J., Yi, P., Tian, D., Chai, Y., Li, W., Guangshuo, O. (2014). A proneural gene controls *C. elegans* neuroblast asymmetric division and migration. FEBS Lett. *588*, 1136–1143.

Zhuang, Y., Barndt, R. J., Pan, L., Kelley, R., & Dai, M. (1998). Functional Replacement of the Mouse E2A Gene with a Human HEB cDNA. *Molecular and Cellular Biology*, *18*(6), 3340–3349. https://doi.org/10.1128/mcb.18.6.3340